

# IOWA STATE UNIVERSITY

## Digital Repository

---

Graduate Theses and Dissertations

Iowa State University Capstones, Theses and  
Dissertations

---

2013

# Fundamental limitations on communication channels with noisy feedback: information flow, capacity and bounds

Chong Li

*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>

 Part of the [Engineering Commons](#)

---

## Recommended Citation

Li, Chong, "Fundamental limitations on communication channels with noisy feedback: information flow, capacity and bounds" (2013). *Graduate Theses and Dissertations*. 13421.  
<https://lib.dr.iastate.edu/etd/13421>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

**Fundamental limitations on communication channels with noisy feedback:  
information flow, capacity and bounds**

by

**Chong Li**

**A dissertation submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of  
DOCTOR OF PHILOSOPHY**

**Major: Electrical Engineering**

**Program of Study Committee:**

**Nicola Elia, Major Professor**

**Zhengdao Wang**

**Aditya Ramamoorthy**

**Aleksandar Dogandžić**

**Wolfgang Kliemann**

**Iowa State University**

**Ames, Iowa**

**2013**

**Copyright © Chong Li, 2013. All rights reserved.**

## DEDICATION

I would like to dedicate this dissertation to my loved family.

## TABLE OF CONTENTS

<b>LIST OF FIGURES</b> . . . . .	vii
<b>ACKNOWLEDGEMENTS</b> . . . . .	x
<b>ABSTRACT</b> . . . . .	xi
<b>CHAPTER 1. INTRODUCTION</b> . . . . .	1
1.1 Background and Motivation . . . . .	1
1.2 Literature Review: Communication Systems with Feedback . . . . .	3
1.2.1 Feedback Control with Communication Constraints . . . . .	3
1.2.2 Feedback Information Theory . . . . .	5
1.3 Thesis Contributions . . . . .	7
1.4 Thesis Outline . . . . .	8
1.5 Notations . . . . .	10
<b>CHAPTER 2. PRELIMINARIES</b> . . . . .	11
2.1 Entropy, Mutual Information and Causal Conditioning Entropy . . . . .	11
2.1.1 Entropy and Mutual Information . . . . .	11
2.1.2 Causal Conditioning Entropy . . . . .	12
2.2 Directed Information . . . . .	13
2.3 Cover-Pombra (CP) Scheme . . . . .	15
2.4 Conclusion . . . . .	17
<b>CHAPTER 3. INFORMATION FLOW ANALYSIS IN COMMUNICATION</b>	
<b>CHANNELS WITH NOISY FEEDBACK</b> . . . . .	19
3.1 Noisy Feedback and Causality . . . . .	19
3.2 Residual Directed Information . . . . .	23

3.3	Information Flow Decomposition in Noisy Feedback Channels . . . . .	26
3.4	Conclusions . . . . .	30
<b>CHAPTER 4. A CHANNEL CODING THEOREM AND BOUNDS ON</b>		
<b>THE CAPACITY OF FINITE-ALPHABET CHANNELS WITH NOISY</b>		
<b>FEEDBACK . . . . .</b>		
4.1	Introduction . . . . .	31
4.2	Problem Formulation and Preliminaries . . . . .	32
4.2.1	Technical Lemmas . . . . .	35
4.3	Channel Coding Theorem . . . . .	36
4.4	Capacity Bounds . . . . .	37
4.5	Case Study: Capacity Bounds for Finite State Channels . . . . .	42
4.5.1	Necessary Technical Lemmas . . . . .	43
4.5.2	Proof of Theorem 39 . . . . .	44
4.6	Conclusions . . . . .	46
<b>CHAPTER 5. BOUNDS ON THE CAPACITY OF GAUSSIAN CHAN-</b>		
<b>NELS WITH NOISY FEEDBACK . . . . .</b>		
5.1	Introduction . . . . .	47
5.2	Information Flow Revisited . . . . .	52
5.3	Upper Bound Characterization . . . . .	56
5.3.1	Upper Bound On The N-block Capacity . . . . .	56
5.3.2	Characterization Under Stationary Gaussian Noises . . . . .	61
5.4	Lower Bound Characterization . . . . .	65
5.4.1	Lower Bound On The N-block Capacity . . . . .	65
5.4.2	Characterization under Stationary Gaussian Channels . . . . .	69
5.5	Simulation Results . . . . .	71
5.6	Conclusion . . . . .	75
<b>CHAPTER 6. CAPACITY-ACHIEVING CODES FOR NOISY FEEDBACK</b>		
<b>CHANNELS: A NECESSARY CONDITION . . . . .</b>		
		76

6.1	Introduction . . . . .	76
6.2	Modeling and Preliminaries . . . . .	77
6.3	Rate-Loss in Using Noisy Feedback . . . . .	79
6.4	An Example: Noisy Feedback AWGN Channels . . . . .	84
6.5	Conclusion . . . . .	86
 <b>CHAPTER 7. NOISY FEEDBACK COMMUNICATIONS WITH SIDE</b>		
<b>INFORMATION AT THE DECODER . . . . .</b>		<b>87</b>
7.1	Introduction . . . . .	87
7.2	Modeling . . . . .	89
7.3	Capacity Characterization . . . . .	90
7.3.1	Capacity of Gaussian Noisy Feedback Channels with Side Information at the Decoder . . . . .	92
7.3.2	Calculation of A Lower Bound: An Iteration Algorithm . . . . .	94
7.4	Simple Coding Strategy . . . . .	97
7.5	Conclusion . . . . .	101
 <b>CHAPTER 8. CONCLUSION AND FUTURE RESEARCH DIRECTIONS</b>		<b>103</b>
8.1	Conclusion . . . . .	103
8.2	Future Directions . . . . .	104
8.2.1	Noisy feedback capacity and computable bounds . . . . .	104
8.2.2	To use or not to use feedback . . . . .	105
8.2.3	Explicit linear feedback coding schemes . . . . .	106
8.2.4	Multi-terminal communications with noisy feedback . . . . .	107
8.2.5	Confluence of feedback control and feedback communication . . . . .	107
 <b>APPENDIX A. PROOFS OF RESULTS IN CHAPTER 3 . . . . .</b>		<b>109</b>
 <b>APPENDIX B. PROOFS OF RESULTS IN CHAPTER 4 . . . . .</b>		<b>112</b>
 <b>APPENDIX C. PROOFS OF RESULTS IN CHAPTER 5 . . . . .</b>		<b>117</b>
 <b>APPENDIX D. PROOFS OF RESULTS IN CHAPTER 6 . . . . .</b>		<b>127</b>

APPENDIX E. PROOFS OF RESULTS IN CHAPTER 7 . . . . .	128
BIBLIOGRAPHY . . . . .	135

## LIST OF FIGURES

Figure 2.1	Gaussian channels with noiseless feedback . . . . .	15
Figure 2.2	Gaussian channels with noiseless feedback: Cover-Pombra scheme . . .	17
Figure 2.3	A concatenated coding representation of the CP scheme. . . . .	18
Figure 3.1	Communication channels with noisy feedback . . . . .	20
Figure 3.2	Family of feedback links in communication systems: the “typical noisy feedback” is the case which we are interested in. . . . .	22
Figure 3.3	Communication Channels with additive noise feedback . . . . .	29
Figure 3.4	The information flow of channels with additive noise feedback . . . . .	29
Figure 4.1	Channels with noisy feedback (a code-function representation) . . . . .	32
Figure 4.2	Finite state channels with noisy feedback . . . . .	42
Figure 5.1	Gaussian channels with additive Gaussian noise feedback . . . . .	48
Figure 5.2	Gaussian channels with additive Gaussian noise feedback(Gaussian sig- nalling and linear feedback) . . . . .	59
Figure 5.3	A concatenated coding representation of CP-like Scheme. The inner linear encoder can be also interpreted as a portion of the equivalent non-feedback channel. . . . .	59
Figure 5.4	New constructed Gaussian channel with noisy feedback . . . . .	67
Figure 5.5	The equivalent Gaussian channel with noiseless feedback . . . . .	69



Figure 5.6	LQG coding scheme: $r \in R^m$ is a vector of white Gaussian noises with zero mean and unit variance without loss of generality, and $H$ is a stable LTI filter. $M$ can be treated as an initial condition of the system $G$ , as described in Elia (2004); Ardestanizadeh and Franceschetti (2012) . . .	71
Figure 5.7	Capacity bounds for MV(1) channel with AWGN feedback. . . . .	72
Figure 5.8	Convergence of the n-block upper bounds ( $\sigma^2 = 0.3, 0.5, 0.7$ ). . . . .	74
Figure 6.1	Communication channels with noisy feedback: feedback information is allowed to be encoded. . . . .	77
Figure 6.2	A concatenated coding scheme for channels with noisy feedback, where $\mathbb{E}_3$ represents the outer encoder and $(\mathbb{E}_1, \mathbb{E}_2)$ represents the inner encoder. 84	
Figure 7.1	Noisy feedback communication channels with side information at the decoder: the feedback information $Z_i$ is known by the decoder after $T$ -step delay. . . . .	88
Figure 7.2	Grasper with noisy force feedback . . . . .	88
Figure 7.3	Additive Gaussian channel with additive Gaussian noise feedback: $W^n \sim \mathbb{N}_n(0, \mathbf{K}_{w,n})$ and $V^n \sim \mathbb{N}_n(0, \mathbf{K}_{v,n})$ where $\mathbf{K}_{w,n}$ and $\mathbf{K}_{v,n}$ are non-singular covariance matrices. Average power constraints $P_1$ and $P_2$ are implemented on inputs $X$ and $U$ , respectively. . . . .	93
Figure 7.4	The n-block achievable rate of a Gaussian noisy feedback channel with side information at the decoder: The forward channel is assumed to be a first-order moving average (1st-MV) Gaussian process, Namely, $W_i = \tilde{U}_i + 0.2\tilde{U}_{i-1}$ where $\tilde{U}_i$ is a white Gaussian process with zero mean and unit variance; The feedback channel is assumed to be an additive white Gaussian noise with variance $\sigma^2$ ; The power constraint is $P_1 = P_2 = 10$ and the coding block length $n = 30$ . . . . .	97
Figure 7.5	1st-MV Gaussian channel with intermittent feedback. . . . .	99
Figure 7.6	Achievable rates of coding scheme $\mathcal{C}$ under power constraints $P_1 = P_2 = 0.5$ . . . . .	101

Figure E.1 Vector representation of the coding scheme (E.1) where  $\mathbf{E}_n = \mathbf{B}_n(\mathbf{I}_n + \mathbf{D}_n\mathbf{B}_n)^{-1}$  and  $\mathbf{F}_n = (\mathbf{I}_n + \mathbf{B}_n\mathbf{D}_n)^{-1}$ . Note that  $\mathbf{E}_n$  is strictly triangular, capturing the one-step feedback delay. . . . . 129

## ACKNOWLEDGEMENTS

I would like to express my thanks to those who supported and helped me with various aspects of conducting my Ph.D research. I would like to especially thank my advisor Professor Nicola Elia for his care and effort to make me become a mature researcher. In these five years, I have been enjoying the inspirational discussion a lot with Professor Elia. His insights on difficult problems and words of encouragement have often inspired me and renewed my hopes for the research. I further thank Professor Zhengdao Wang for his knowledgable guidance and helpful discussions during my Ph.D study. His technical idea and suggestions always refreshed me. I would also like to thank my committee members for their efforts and contributions to this work: Dr. Aditya Ramamoorthy, Dr. Aleksandar Dogandžić and Dr. Wolfgang Kliemann.

The atmosphere here at Iowa State University is great and unforgettable. Many people contributed to the charm of my five-year life here at Ames. My officemates Dr. Jing Wang, Dr. Andalam Satya Mohan Vamsi, Kai Wang, Dr. Diwadkar Amit Vivek shared my either joyful or frustrated moments in my research. Many friends out of my research group made my every single day at Ames. I am grateful for their support, friendship and encouragement and I wish all the bests for them. I thank my colleagues at Qualcomm R&D at New Jersey, who contributed to my great harvest internship.

The final word of my gratitude must go to my closest loved ones, my family members, for their (long-distance) listen, advice and support, without whom I could not make such a success in my study and career.

## ABSTRACT

Since the success of obtaining the capacity (i.e. the maximal achievable transmission rate under which the message can be recovered with arbitrarily small probability of error) for non-feedback point-to-point communication channels by C. Shannon (in 1948), Information Theory has been proved to be a powerful tool to derive fundamental limitations in communication systems. During the last decade, motivated by the emerging of networked systems, information theorists have turned lots of their attention to communication channels with feedback (through another channel from receiver to transmitter). Under the assumption that the feedback channel is noiseless, a large body of notable results have been derived, although much work still needs to be done. However, when this ideal assumption is removed, i.e., the feedback channel is noisy, only few valuable results can be found in the literature and many challenging problems are still open.

This thesis aims to address some of these long-standing noisy feedback problems, with concentration on the *channel capacity*. First of all, we analyze the fundamental information flow in noisy feedback channels. We introduce a new notion, the *residual directed information*, in order to characterize the noisy feedback channel capacity for which the standard directed information can not be used. As an illustration, finite-alphabet noisy feedback channels have been studied in details. Next, we provide an information flow decomposition equality which serves as a foundation of other novel results in this thesis.

With the result of information flow decomposition in hand, we next investigate time-varying Gaussian channels with additive Gaussian noise feedback. Following the notable Cover-Pombra results in 1989, we define the n-block noisy feedback capacity and derive a pair of n-block upper and lower bounds on the n-block noisy feedback capacity. These bounds can be obtained by efficiently solving convex optimization problems. Under the assumption of stationarity on the additive Gaussian noises, we show that the limits of these n-block bounds can be characterized

in a power spectral optimization form. In addition, two computable lower bounds are derived for the Shannon capacity.

Next, we consider a class of channels where feedback could not increase the capacity and thus the noisy feedback capacity equals to the non-feedback capacity. We derive a necessary condition (characterized by the *directed information*) for the capacity-achieving channel codes. The condition implies that using noisy feedback is detrimental to achievable rate, i.e., the capacity can not be achieved by using noisy feedback.

Finally, we introduce a new framework of communication channels with noisy feedback where the feedback information received by the transmitter is also available to the decoder with some finite delays. We investigate the capacity and linear coding schemes for this extended noisy feedback channels.

To summarize, this thesis firstly provides a foundation (i.e. information flow analysis) for analyzing communications channels with noisy feedback. In light of this analysis, we next present a sequence of novel results, e.g. channel coding theorem, capacity bounds, etc., which result in a significant step forward to address the long-standing noisy feedback problem.

## CHAPTER 1. INTRODUCTION

### 1.1 Background and Motivation

Nowadays the widespread availability of large communication networks is allowing unprecedented interactions among the communicating nodes and pointing at interactive communications systems. These systems constantly affect each other via exchanging information over the available communication links. This dynamic, interactive aspect of communication is common to many networked systems, from social networks to biological networks. Unfortunately the success of Information Theory has mostly pertained to unidirectional point-to-point communication systems and most of the traditional information theory results do not help in addressing these interacting networked problems, since they do not handle the causality and real-time constraints of networked systems. Clearly, these new problems require a multidisciplinary approach, which relies on new information theory concepts, utilizes the rich knowledge of the control of uncertain system and adopts the new advances in optimization methods. One key problem, which is the focus of this thesis, is the study of simple interacting communication systems where two systems, the encoder and the decoder, exchange information over noisy forward and feedback channels.

Unfortunately, the current feedback information theory mostly assumes a noiseless feedback channel. Using noiseless feedback provides a plenty of benefits as summarized below:

1. Increase the capacity of communications channels with memory;
2. Significantly simplify the encoding/decoding structure;
3. Improve the decaying rate of the probability of decoding error.

However, the assumption of noiseless feedback in communication systems has long been recognized as the Achilles heel of the information-theoretic study of feedback (Draper and Sahai (2008)). Lucky (1973) stated it:

*feedback communications was an area of intense activity in 1968... A number of authors had shown constructive, even simple, schemes using noiseless feedback to achieve Shannon-like behavior... The situation in 1973 is dramatically different... The subject itself seems to be a burned out case... In extending the simple noiseless feedback model to allow for more realistic situations, such as noisy feedback channels, bandlimited channels, and peak power constraints, theorists discovered a certain brittleness or sensitivity in their previous results.*

Up to present, the “noisy feedback” problem is still open and is considered as a bottleneck in the development of Information Theory. As we known, in most interacting networked systems, the feedback channel is inevitably noisy. To get over the feedback noise such that the theoretical “noiseless” assumption holds, in current industry applications, as an example, the intensive error-correcting code is widely used in the feedback channel. This implementation definitely requires high transmission power and allows limited transmission rate in the feedback. The lack of a mathematical theory in noisy feedback communications is considered to be an obstruction in further industrial development in the field of wireless communications. By using noisy feedback, do we still have the aforementioned benefits? If yes, how much can we obtain from it? If no, why?

Besides the motivation arising from the development of Information Theory, the incomplete unified theory of feedback control and communications with feedback triggers the thesis as well. In this decade, some pioneer researchers have successfully applied control system ideas and results to develop analysis and design tools for communication systems with noiseless feedback. See Elia (2004); Liu et al. (2004a,b); Liu and Elia (2005, 2006) and reference therein. The approaches carried out from a control theory perspective provide a novel avenue (compared with the approaches produced by information theorists) to discover the fundamental benefits of using feedback. As an illustration, Elia (2004) has shown the equivalence between feedback

stabilization over a communication channel and communication with noiseless feedback. One impressive consequence of this equivalence is that we can use many results and controller design methods from control theory to analyze and design communication systems with noiseless feedback. To the best knowledge of mine, however, no literature has extended this work to the noisy feedback case due to many theoretic difficulties. One main difficulty is due to the non-classical information pattern, addressed by Witsenhausen (1968) in his famous counterexample. In particular, it is due to the loss of coordination between the feed-forward encoder/controller and the feedback encoder/controller in the noisy feedback system.

## **1.2 Literature Review: Communication Systems with Feedback**

The literature review will proceed from two aspects where most of the relevant work has been done by two distinguished group of researchers in the last decade. Both of these aspects trigger the work on communication systems with noisy feedback in this thesis.

### **1.2.1 Feedback Control with Communication Constraints**

Researchers in control theory community have been doing the study of problems connected to the presence of communication channels in feedback control systems. Wong and Brockett (1999); Tatikonda and Mitter (2000); Tatikonda (2000); Nair and Evans (2000); Matveev and Savkin (2001); Sahai (2001); Elia and Mitter (2001); Elia (2002, 2003); Baillieul (2002) are a few of the earlier publications in this area. In the study of the interaction and integration of communication and control, important questions pertain to the benefits of feedback in communication systems. As the literature is vast, in what follows, we have a brief discussion on one research direction which is related to the results in this thesis. This direction invokes the idea of viewing the feedback communication system as a control system and then utilizes the mature control theory to address many long-standing problems in feedback communication. As a well-known pioneer work in this direction, Tatikonda (2000); Tatikonda and Mitter (2009) proposed a unified view of feedback control and feedback information theory. Specifically, they extended Dobrushin's idea of treating the feedback communication systems as inter-connections



of stochastic kernels, and then proved that the feedback capacity of communication channels with noiseless feedback can be characterized as the supremum of directed information rate (introduced by Massey (1990)) from the channel inputs to the channel outputs. In addition, Tatikonda and Mitter (2009) reformulated the optimization problem of computing the multi-letter feedback capacity as a stochastic control problem, and developed a dynamic programming solution to compute the finite-horizon feedback capacity.

Sahai (2001, 2006) observed the insufficiency of Shannon capacity while communicating delay-sensitive information streams over general noisy communication channels. The information streams may include non-stationary, non-ergodic sources. Motivated by this observation, a fundamental theory, *anytime information theory*, was proposed, which is closely related to the control problem of tracking unstable sources over noisy channels. Briefly speaking, anytime capacity, corresponding to moment stability of the associated control systems, is stronger than the Shannon capacity, corresponding to almost sure stability of the associated control system. For Gaussian channels, anytime capacity equals the Shannon capacity since the moment stability is equivalent to almost sure stability.

Elia (2004) established the general equivalence between reliable feedback communication and feedback stabilization over Gaussian channels with noiseless feedback. By taking additive white Gaussian noise(AWGN) channel into account, the celebrated Schalkwijk-Kailath(SK) coding scheme (for reliable communication) is nothing but a rewrite of the stabilization of a special linear quadratic Gaussian (LQG) system. In particular, Elia proved that the transmission rate over the channel (characterized by the directed information) equals the degree of instability of the open-loop system (characterized by the unstable eigenvalues of the systems).

Yang et al. (2005) applied the stochastic control formulation to compute the feedback capacity for a discrete-input finite-state Markov channel, which is characterized by the directed information rate. They also obtained the optimal input distribution to the channel, which has the Markov property, and thus reduced the infinite-horizon stochastic control optimization problem to a tractable one.

Motivated by the work listed above, Liu (2006) addressed the problem of identifying the limits of Gaussian channels with noiseless feedback from a unified perspective. In particu-

lar, Liu and Elia established a general equivalence among feedback communication, estimation, and feedback stabilization over the feedback Gaussian channels. *The achievable communication rates in the feedback communication problems can be alternatively obtained by the decay rates of the Cramer-Rao bounds in the associated estimator problems or by the Bode sensitivity integrals in the associated control problem.* In light of this fundamental equivalence, they showed the optimality of the Kalman filtering algorithm in feedback communication, estimation, and feedback control. We refer the interested readers to Liu et al. (2004b,a); Liu and Elia (2005) for more details.

### 1.2.2 Feedback Information Theory

Information theorists since Shannon have been interested in the effect of feedback on the theoretical transmission rate limits achievable by a communication channel, e.g. Schalkwijk and Kailath (1966); Schalkwijk (1966, 1968); Cover and Pombra (1989); Pombra and Cover (1994); Butman (1969); Ozarow (1984); Ozarow and Leung (1984); Kramer (2002b); Shahar-Doron and Feder (2004); Yang (2004, 2007); Permuter et al. (2009); Shayevitz and Feder (2011). The literature review in this section focuses on the point-to-point communications with feedback as the multi-terminal case is out of the scope of this thesis.

Shannon (1958) first looked into communications with feedback, and proved a notable result that feedback could not increase capacity of discrete memoryless channels (DMCs). However, by explicitly proposing a feedback communication scheme (e.g. Schalkwijk and Kailath (1966); Schalkwijk (1966)), it is found that feedback can greatly decrease the complexity of encoders/decoders and improve other performance measures like the decay of the probability of decoding errors.

As a sequential work, Butman (1969) showed that feedback increases capacity for the first-order autoregressive channels. Then Tiernan and Schalkwijk (1974) provided upper bounds on the capacity of band-limited first-order Gaussian autoregressive channels with noiseless feedback under an average energy constrain. Butman (1976) achieved tighter bounds on the capacity of general  $m$ -th order gaussian autoregressive channels with linear feedback. Motivated by these notable work on memory Gaussian channels, Cover and Pombra (1989) proposed

an  $n$ -block capacity on Gaussian channels with noiseless feedback where the additive noise is assumed to be *time-varying/arbitrary* Gaussian. In the same paper, they characterized this  $n$ -block feedback capacity in a matrix optimization form. Vandenberghe et al. (1998) showed that this matrix optimization problem can be transformed into a convex form and efficiently solved by semi-definite programming (SDP). Moreover, based on the capacity characterization, Cover and Pombra (1989) showed that the feedback could not increase the capacity by half bit, which is a complementary result to the one derived by Ebert (1970), saying, feedback could not increase the capacity by factor two compared with the non-feedback capacity. This result was further refined in Dembo (1989) and Chen and Yanagi (1999). Although the  $n$ -block feedback capacity is explicitly characterized, evaluate its asymptotic value (i.e. block length goes to infinity) is notoriously difficult. Kim (2010) characterized this asymptotic value in the form of power spectral optimization by assuming the stationarity on Gaussian noise. However, this single infinite dimensional optimization problem is still difficult to solve except the first-order autoregress moving averaging (i.e. ARMA(1)) channel. We refer to Elias (1956, 1967); Wolfowitz (1975); Ihara (1980, 1988); Ozarow (1990a,b); Ihara (1990, 1994) and reference therein for other related work on feedback Gaussian channels.

Although it was not clearly mentioned in the above feedback Gaussian literature, the directed information proposed by Massey (1990) is the measure for the feedback capacity and the feedback Gaussian capacity is nothing but the characterization on the directed information. In light of the directed information, researchers turned to investigate other interesting channels and characterized their capacities. We list a few of them as follows. Kim (2008) provided a channel coding theorem for feedback channel with finite memory, in other words, the current channel output is the function of the current and past  $m$  symbols from the channel inputs and the stationary ergodic channel noise. Tatikonda and Mitter (2009) provide a channel coding theorem for finite-alphabet channel with arbitrary noises by extending Verdú and Han's work on the general non-feedback channel capacity (Verdú and Han (1994)). Permuter et al. (2009) investigated the capacity of finite state channels with time-invariant deterministic feedback by extending Gallager's idea of characterizing the capacity of finite state non-feedback channels (Gallager (1968)). We remark, again, that all these results were obtained by characterizing the

directed information for corresponding channels with *noiseless* feedback.

As it is shown above, all these work assumed that the feedback is noiseless, which is not the case in most practical scenarios. Unfortunately, until present, only few papers have studied communication channels with noisy feedback. We briefly classify the results in the literature into two main categories. The first category studies the usefulness of noisy feedback by investigating reliability functions and error exponents. See Draper and Sahai (2006b); Kim et al. (2007); Burnashev and Yamamoto (2008). The second category (Omura (1968); Lavenberg (1971); Martins and Weissman (2008); Chance and Love (2010); Kumar et al. (2009), Zhang and Guo (2011)) focuses on the derivation of coding schemes mostly for additive Gaussian channels with noisy feedback based on the well-known SK scheme (Schalkwijk and Kailath (1966)). Motivated by these few fragmented results in the literature, we herein wish to provide a comprehensive mathematical theory of communication channels with noisy feedback.

### 1.3 Thesis Contributions

The main results of the thesis are summarized as follows.

1. To comprehensively analyze/understanding the noisy feedback problem, we first investigate the information flow in noisy feedback channels. We propose a new concept, *residual directed information*, in order to capture the capacity of noisy feedback channels. Next, we derive a fundamental equality, *information flow decomposition equality*, as a basis of all the other results in the thesis.
2. As the first application of the new concept and the information flow equality, we provide a channel coding theorem and bounds for the capacity of finite-alphabet communication channels with noisy feedback. Then we consider a specific class of channels, finite-state finite-alphabet channels, and provide upper bounds on the capacity.
3. We study time-varying Gaussian channels with additive Gaussian noise feedback. We extend the well-known result of Cover-Pombra on noiseless feedback Gaussian channels to noisy feedback settings. First of all, we define the *n-block noisy feedback capacity* and then derive a pair of n-block upper and lower bounds on the n-block capacity. These bounds

can be obtained by solving convex optimization problems. Next, under the assumption that the additive Gaussian noises are stationary, we prove that the limits of the upper and lower bounds exist and can be characterized in a form of power spectral optimization, thus providing bounds on the asymptotic Shannon capacity. Finally, two computable lower bounds on Shannon capacity are provided.

4. We provide a necessary condition on the capacity-achieving channel codes of noisy feedback channels. For a special class (e.g. DMC) of channels where the noisy feedback capacity is equal to the non-feedback capacity, this condition implies that using noisy feedback is detrimental to achievable rate, i.e., could not achieve the capacity.
5. As an initial looking-forward work, we consider a new framework of communication channels with noisy feedback where the feedback information received by the encoder is also available to the decoder with some finite delays. We first show that the feedback capacity can be characterized in terms of the causal conditioning directed information. Then we propose a specific linear coding scheme with good transmission rate for Gaussian channels under the new framework.

We remark that the first result of information flow analysis is the root of the thesis and induces all the rest of results mentioned above.

## 1.4 Thesis Outline

Chapter 1 provides the background and motivations of the study on communication channels with noisy feedback. In addition, we review the relevant literature on networked control systems with communication constraints and information theory of communications with feedback, and outline the thesis.

Chapter 2 presents relevant preliminary results on communications with noiseless feedback, which provide hints and serve as useful tools to address the noisy feedback problem.

Chapter 3 studies the information flow in communication channels with noisy feedback. Different from the non-feedback and noiseless feedback settings, the directed information flowing from the channel inputs to the channel outputs does not deliver the message at full usage. We

derive an information flow decomposition equality which reveals the information flow pattern in noisy feedback channels. Namely, there exist three information flows in the forward channel: message-delivery flow, feedback-noise-delivery flow, and the interference flow between these two flows. In addition, we propose a new concept, residual directed information, to capture the message-delivery flow in a compact form.

Chapter 4 considers finite-alphabet communication channels with noisy feedback. We extend the idea of Tatikonda and Mitter, characterizing the capacity of finite-alphabet channels with noiseless feedback, to our noisy feedback settings, and prove a channel coding theorem characterized by the residual directed information. We next investigate the finite-state channels with noisy feedback and provide an upper bound on the capacity. Chapter 5 studies Gaussian channels with additive Gaussian noise feedback. We begin with arbitrary additive Gaussian noises and define the  $n$ -block noisy feedback capacity, which has operational meaning as  $n \rightarrow \infty$ . In light of the information flow analysis, we derive an upper bound on the  $n$ -block capacity which can be characterized in a convex form and then can be obtained by using standard technical optimization tools. As a counterpart, we use a novel approach to derive an lower bound on the  $n$ -block capacity which is characterized in a convex form as well. This lower bound is not restricted to any specific coding scheme and holds for any additive Gaussian noises. In order to find bounds on the Shannon capacity defined in asymptotic fashion, we assume that the additive Gaussian noises are stationary. We then prove the limits of the upper and lower bounds exist and can be characterized in a form of power spectral optimization. Finally, two approaches of computing lower bounds on the Shannon capacity are provided.

Chapter 6 provides a necessary condition on the capacity-achieving channel codes of noisy feedback channels, by utilizing the information flow analysis. As it is known, although feedback increases channel capacity in general, it does not for certain class of channels, e.g., discrete memoryless channels (DMCs). We denote this class of channels as *feedback-unfavorable* channels. Then the derived necessary condition for feedback-unfavorable channels indicates that any capacity-achieving channel code has to discard feedback information in order to use channel at its full capacity.

Chapter 7 introduces a new framework of communication channels with noisy feedback

where the feedback information received by the transmitter is also available to the decoder with some finite delays. The merits of this new framework are demonstrated by two aspects: 1) Its capacity can be characterized by the causal conditioning directed information. As an illustration, we characterize the  $n$ -block capacity of Gaussian noisy feedback channels under our new framework and propose an iteration algorithm to obtain a lower bound; 2) By constructing a specific linear coding scheme for the first-order moving average Gaussian channels with intermittent(erasure) feedback, we show that the new framework allows linear feedback coding schemes with positive transmission rate, which is (in certain regime) much larger than the non-feedback Gaussian channel capacity.

Chapter 8 concludes the thesis and provides some avenues for future research.

## 1.5 Notations

Uppercase and corresponding lowercase letters (*e.g.*  $Y, Z, y, z$ ) denote random variables and realizations, respectively. The probability distribution of random variables is denoted by only  $p$  when the arguments of  $p$  specify the distribution. For example, the value  $p_{XY}(x, y)$  of the joint distribution  $p_{XY}$  of the random variables  $X$  and  $Y$  is written simply as  $p(x, y)$ .

We consider only positive integer subscripts for symbols. For simplicity of notation, we sometimes allow non-positive subscripts, which refers to an empty symbol. For example,  $X^n Y^{n-1}$  for  $n = 1$  is equivalent to  $X^1$  as  $Y^0 = \emptyset$ .

$x^n$  represents the vector  $[x_1, x_2, \dots, x_n]^T$  and  $x^0 = \emptyset$ .  $\mathbf{I}_n$  represents an  $n \times n$  identity matrix.  $\mathbf{K}_n \succ 0$  ( $\mathbf{K}_n \succeq 0$ ) denotes that the  $n \times n$  matrix  $\mathbf{K}_n$  is positive definite (semi-definite).  $\log$  denotes the logarithm base 2 and  $0 \log 0 = 0$ . The expectation operator over  $X$  is presented as  $\mathbb{E}(X)$ .

## CHAPTER 2. PRELIMINARIES

In this chapter we present some mathematical preliminaries which we will be using enroute to deriving our results.

### 2.1 Entropy, Mutual Information and Causal Conditioning Entropy

#### 2.1.1 Entropy and Mutual Information

We first introduce a well known concept, *entropy*, which is a measure of the uncertainty of a random variable.

**Definition 1** (*Entropy*)(Cover and Thomas (2006)) Let  $X$  be a discrete random variable with alphabet  $\mathcal{X}$  and probability mass function  $p(x) = \Pr(X = x), x \in \mathcal{X}$ . The entropy  $H(X)$  is defined by

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x),$$

and  $H(X) \geq 0$ .

If  $X$  is a continuous random variable, we have another concept of *differential entropy*.

**Definition 2** (*Differential Entropy*) Let  $X$  be a continuous random variable with support set  $\mathcal{S}$  and probability density function is denoted by  $f(x)$ . The differential entropy  $h(X)$  is defined by

$$h(X) = - \int_{\mathcal{S}} f(x) \log f(x) dx.$$

We now recall a useful lemma on the differential entropy as follows.

**Lemma 3** Let a random vector  $X^n \in \mathbb{R}^n$  have zero mean and covariance  $\mathbf{K}_{x,n} = \mathbb{E}X^n X^{nT}$  (i.e.  $\mathbf{K}_{x,n}(i, j) = \mathbb{E}X_i X_j, 1 \leq i, j \leq n$ ). Then

$$h(X^n) \leq \frac{1}{2} \log(2\pi e)^n \det \mathbf{K}_{x,n}$$



with equality if and only if  $X^n \sim \mathcal{N}(0, \mathbf{K}_{x,n})$ .

**Definition 4** The mutual information  $I(X; Y)$  between two discrete random variables with a joint probability mass function  $p(x, y)$  and marginal probability mass function  $p(x)$  and  $p(y)$  is defined by

$$I(X; Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy = H(Y) - H(Y|X),$$

and the mutual information density is defined by

$$i(X; Y) = \log \frac{p(X, Y)}{p(X)p(Y)}$$

Similarly, the mutual information  $I(X; Y)$  between two continuous random variables with joint density  $f(x, y)$  is defined as

$$I(X; Y) = \int f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dx dy = h(Y) - h(Y|X).$$

and the mutual information density is defined by

$$i(X; Y) = \log \frac{f(x, y)}{f(x)f(y)}.$$

The mutual information is of ultimate importance in information theory. It is a measure of the amount of information that one random variable contains about the other random variable. As it is shown, it can be also interpreted as the reduction of the uncertainty of one random variable given the knowledge of the other.

In the following context of this section, we restrict ourself to the definitions/results of discrete random variables as the definitions/results of continuous random variables directly follow in parallel.

### 2.1.2 Causal Conditioning Entropy

Consider a time-ordered random variable sequence  $(X^n, Y^n)$  as follows,

$$X_1, Y_1, X_2, Y_2, \dots, X_{n-1}, Y_{n-1}, X_n, Y_n. \quad (2.1)$$

We first define the causal conditioning probability as

$$p(y^n || x^n) = \prod_{i=1}^n p(y_i | x^i, y^{i-1}).$$

**Definition 5** *The entropy of a sequence of discrete random variables  $Y^n$ , causally conditioning on a sequence of discrete random variables  $X^n$  is defined by*

$$H(Y^n||X^n) = \mathbb{E}_{p(X^n, Y^n)} \log p(Y^n||X^n) = \sum_{i=1}^n H(Y_i|Y^{i-1}, X^i)$$

Notice that the only difference between the causal conditioning entropy and the classical entropy is the replacement of  $X^n$  by  $X^i$ . The term “causal” reflects the fact that the current random variable  $Y_i$  depends on the past and current  $X^i$ , instead of the whole sequence  $X^n$ .

Now we consider a time-ordered random variable sequence  $(X^n, Y^n, Z^n)$  as follows,

$$X_1, Y_1, Z_1, X_2, Y_2, Z_2, \dots, X_{n-1}, Y_{n-1}, Z_{n-1}, X_n, Y_n, Z_n. \quad (2.2)$$

A vector-valued causal conditioning entropy is defined accordingly as

$$H(Z^n||X^n, Y^n) = \mathbb{E}_{p(X^n, Y^n, Z^n)} \log p(Z^n||X^n, Y^n) = \sum_{i=1}^n H(Z_i|Z^{i-1}, X^i, Y^i),$$

and

$$H(Y^n, Z^n||X^n) = \mathbb{E}_{p(X^n, Y^n, Z^n)} \log p(Y^n, Z^n||X^n) = \sum_{i=1}^n H(Y_i, Z_i|Y^{i-1}, Z^{i-1}, X^i).$$

Now, we consider the “mix” of the causal conditioning and usual conditioning. We herein adopt the notation introduced in Kramer (2002a). Specifically, we use the notational convention that conditioning is done from left to right. Thus, for the time-ordered random variable sequence (2.2),

$$H(Y^n||X^n|Z^n) = \sum_{i=1}^n H(Y_i|Y^{i-1}, X^i, Z^n).$$

## 2.2 Directed Information

To deal with the causality of the system with feedback, Massey (1990) proposed the definition of *directed information* as follows.

**Definition 6** *The directed information from a random variable sequence  $X^n$  to a random variable sequence  $Y^n$  is defined by*

$$I(X^n \rightarrow Y^n) = \sum_{i=1}^n I(X^i; Y_i|Y^{i-1}) = \mathbb{E}_{p(X^n, Y^n)} \log \frac{p(Y^n||X^n)}{p(Y^n)}.$$

Equivalently,

$$I(X^n \rightarrow Y^n) = H(Y^n) - H(Y^n || X^n).$$

The directed information density is defined by

$$i(X^n \rightarrow Y^n) = \log \frac{p(Y^n || X^n)}{p(Y^n)}.$$

We would like to remark that Massey's definition of directed information implicitly restricts the time ordering of random variables  $(X^n, Y^n)$  as (2.1). We refer the interested readers to Tatikonda and Mitter (2009) for the definition of Directed Information for an arbitrary time ordering of random variables.

**Remark 7** *The directed information is of importance in characterizing the capacity of channels with perfect feedback (Kim (2008), Kim (2010), Tatikonda and Mitter (2009)) or deterministic feedback (Permuter et al. (2009)). Moreover, it has valuable interpretation in portfolio theory, data compression and hypothesis testing (Permuter et al. (2011)). However, as we will show in Chapter 3, it fails to characterize the capacity of channels with noisy feedback and is not a proper quantity to work on while analyzing the noisy feedback systems. In addition, directed information is also relevant in a rate distortion problem. Based on the work of Weissman and Merhav (2003) and Pradhan (2004), Venkataramanan and Pradhan (2007) formulated a problem of source coding with feed-forward and showed that directed information can be used to characterize the rate-distortion function. Another source coding problem in which directed information has arisen is investigated by Zamir et al. (2008). In their paper, a linear prediction representation of the rate distortion function of a stationary Gaussian source is captured by directed information.*

We next recall the definition of causal conditioning directed information.

**Definition 8** *The directed information from a random variable sequence  $X^n$  to a random variable sequence  $Y^n$ , causal conditioning on a random variable sequence  $Z^n$ , is defined by*

$$I(X^n \rightarrow Y^n || Z^{n-1}) = \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}, Z^{i-1}) = \mathbb{E}_{p(X^n, Y^n, Z^n)} \log \frac{p(Y^n || X^n, Z^n)}{p(Y^n || Z^n)}.$$

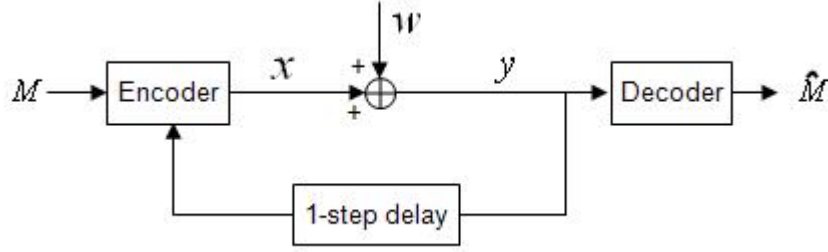


Figure 2.1 Gaussian channels with noiseless feedback

Equivalently,

$$I(X^n \rightarrow Y^n || Z^{n-1}) = H(Y^n || Z^{n-1}) - H(Y^n || X^n, Z^{n-1}).$$

The causal conditional directed information density is defined by

$$i(X^n \rightarrow Y^n || Z^{n-1}) = \log \frac{p(Y^n || X^n, Z^n)}{p(Y^n || Z^n)}$$

Here the underlying time order of the random variable sequence is restricted to the one presented in (2.2).

Furthermore, consider the random variable sequence (2.2) with random variables  $S^n$  given in prior, we define

$$I(X^n \rightarrow Y^n || Z^{n-1} | S^n) = \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}, Z^{i-1}, S^n) = H(Y^n || Z^{n-1} | S^n) - H(Y^n || X^n, Z^{n-1} | S^n).$$

### 2.3 Cover-Pombra (CP) Scheme

We next recall the Cover-Pombra scheme which will be used and extended in several chapters. Consider a discrete-time Gaussian channel with noiseless feedback as shown in Fig.2.1. The additive Gaussian channel is modeled as

$$Y_i = X_i + W_i \quad i = 1, 2, \dots$$

where the gaussian noise  $\{W_i\}_{i=1}^\infty$  satisfies  $W^n = \{W_1, W_2, \dots, W_n\} \sim \mathbb{N}_n(0, \mathbf{K}_{w,n})$  for all  $n \in \mathbf{Z}^+$ . For a code of rate  $R_n$  and length  $n$ , we specify a  $(n, 2^{nR_n})$  channel code as follows.

$M$  is an uniformly distributed message index where  $M \in \{1, 2, 3, \dots, 2^{nR_n}\}$ . There exists an encoding process  $X_i(M, Y^{i-1})$  for  $i = 1, 2, \dots, n$  ( $X_1(M, Y^0) = X_1(M)$ ), with power constraint

$$\frac{1}{n} \sum_{i=1}^n \mathbf{E} X_i^2(M, Y^{i-1}) \leq P,$$

and a decoding function  $g: Y^n \rightarrow \{1, 2, \dots, 2^{nR_n}\}$  with an error probability satisfying

$$P_e^{(n)} = \frac{1}{2^{nR_n}} \sum_{M=1}^{2^{nR_n}} p(M \neq g(y^n) | M) \leq \epsilon_n$$

where  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ . Notice that we do not assume stationarity on  $W$ . Therefore, the classical shannon capacity (i.e. the supremum of all achievable rates  $R$ ) of this feedback Gaussian channel may not exist. Because of this fact, Cover and Pombra in 1989 defined the *n-block feedback capacity* as follows.

**Definition 9** Let  $\{W_i\}_{i=1}^\infty$  be an arbitrary Gaussian stochastic process such that  $W^n \sim \mathbb{N}_n(0, \mathbf{K}_{w,n})$ .

Then  $\{C_{fb,n}^{noisy}\}_{n=1}^\infty$  is a sequence of *n-block noisy feedback capacity* if it satisfies,

1. there exists a sequence of  $(n, 2^{n(C_{fb,n}^{noisy} - \epsilon)})$  noisy feedback channel codes with  $P_e^{(n)} \rightarrow 0$ , as  $n \rightarrow \infty$ , for  $\epsilon > 0$ ;
2. conversely, for  $\epsilon > 0$ , any sequence of  $(n, 2^{n(C_{fb,n}^{noisy} + \epsilon)})$  codes has  $P_e^{(n)}$  bounded away from zero for all  $n$ .

In the same paper, they proposed a coding scheme, consisting of linear encoding of the feedback information and Gaussian signalling of the message, to characterize the n-block capacity  $C_{fb,n}$ . See Fig.2.2. Specifically,

*The channel input signal:*  $X^n = S^n + \mathbf{B}_n W^n$

*The channel output signal:*  $Y^n = S^n + \mathbf{B}_n W^n + W^n$

*The power constraint:*  $\text{tr}(\mathbf{K}_{s,n} + \mathbf{B}_n \mathbf{K}_{w,n} \mathbf{B}_n^T) \leq nP$

where  $S^n \sim N(0, \mathbf{K}_{s,n})$  is the Gaussian-signaled message information vector and  $\mathbf{B}_n$  is an  $n \times n$  strictly lower triangular linear encoding matrix. Note that the one-step delay in the feedback link is captured by the particular structure of matrix  $\mathbf{B}_n$ . Random variables  $S^n$

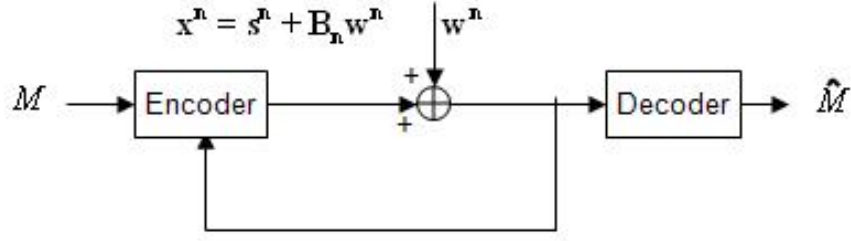


Figure 2.2 Gaussian channels with noiseless feedback: Cover-Pombra scheme

and  $W^n$  are assumed to be independent. This proposed coding scheme can be specifically expressed as a concatenated coding scheme as shown in Fig. 2.3. The outer encoder  $\mathbf{E}_1$  maps each message index to a vector  $s^n$  which is drawn from the distribution  $N(0, \mathbf{K}_{s,n})$ . The inner encoder linearly takes the message information vector and the feedback information to produce channel inputs.

Then, it is proved that, without loss of optimality, the n-block feedback capacity  $C_{fb,n}$  can be characterized by the CP scheme as follows,

$$C_{fb,n} = \max_{\mathbf{B}_n, \mathbf{K}_{s,n}} \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n) \mathbf{K}_{w,n} (\mathbf{I}_n + \mathbf{B}_n)^T + \mathbf{K}_{s,n})}{\det \mathbf{K}_{w,n}} \quad (2.3)$$

where the maximum is taken over all positive semidefinite matrices  $\mathbf{K}_{s,n}$  and all strictly lower triangular matrices  $\mathbf{B}_n$  satisfying

$$\frac{1}{n} \text{tr}(\mathbf{K}_{s,n} + \mathbf{B}_n \mathbf{K}_{w,n} \mathbf{B}_n^T) \leq P. \quad (2.4)$$

## 2.4 Conclusion

In this chapter, we reviewed some information concepts capturing real-time causality in the feedback system, and presented the well-known CP coding scheme used for characterizing the n-block capacity of time-varying noiseless feedback Gaussian channels. These reviewed materials will be frequently used throughout the thesis.

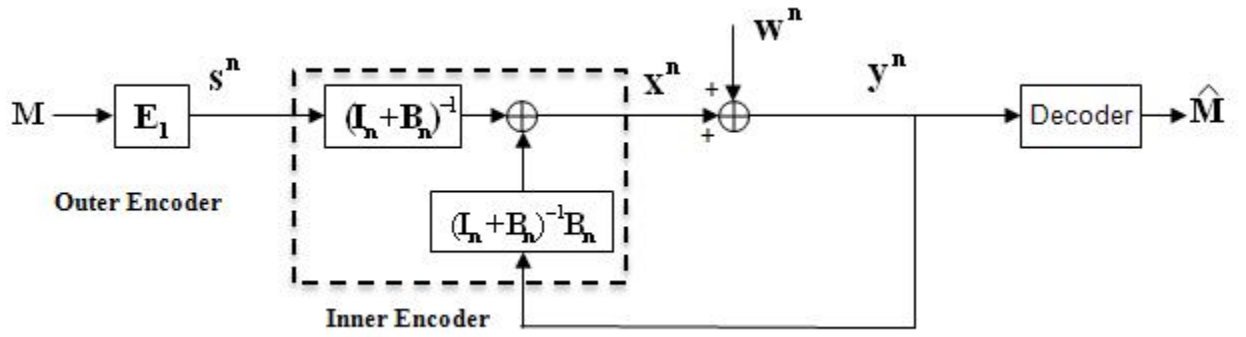


Figure 2.3 A concatenated coding representation of the CP scheme. .

## CHAPTER 3. INFORMATION FLOW ANALYSIS IN COMMUNICATION CHANNELS WITH NOISY FEEDBACK

In this chapter, we analyze the information flow in noisy feedback systems, which provides a foundation for addressing noisy feedback problems Li and Elia (2011b). We first introduce a generic noisy feedback setup considered in this thesis and give a high-level discussion on the failure of using either mutual information or directed information as a measure of the information flow from the transmitter to the receiver, named *effective information flow*, through the communication channel with noisy feedback. Then we define a new measure, named *residual directed information*, and derive its properties. Finally, we provided the information flow decomposition equality which explicitly unveils the distinctive information flow pattern in the noisy feedback channels. As it will be shown, this decomposition equality plays a core role in deriving novel results in the rest of this thesis.

### 3.1 Noisy Feedback and Causality

According to Fig.3.1, we model the channel at time  $i$  as  $p(y_i|x^i, y^{i-1})$ ; namely, the current forward channel output depends on all the previous forward channel outputs, and the previous and current forward channel inputs. The channel output (without any encoding) is then fed back to the encoder through another noisy channel (i.e. feedback link), which is modeled as  $p(z_i|y^i, z^{i-1})$ ; namely, the current feedback link output depends on all the previous feedback link outputs, and the previous and current feedback link inputs. Note that, in order to distinguish the forward channel and the feedback channel, we use *feedback link* to refer to the feedback channel in the sequel of the thesis. At time  $i$ , the deterministic encoder takes the message index  $M$  and the past outputs  $Z_1, Z_2, \dots, Z_{i-1}$  of the feedback link, and then produces a channel



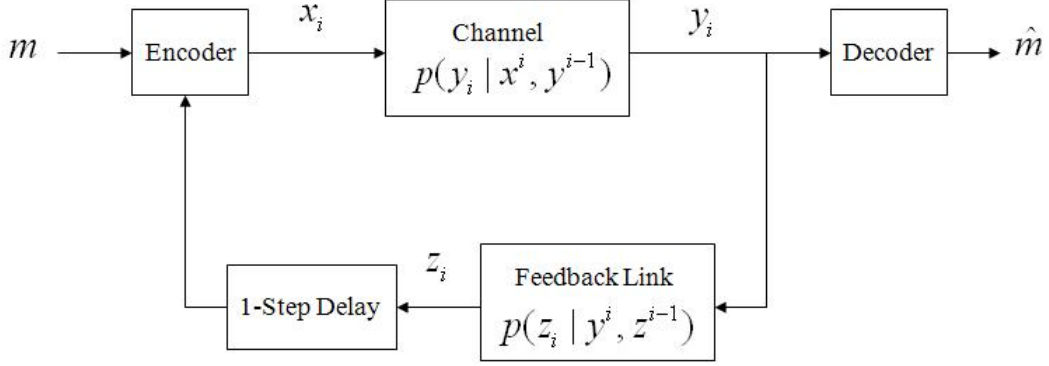


Figure 3.1 Communication channels with noisy feedback

input  $X_i$ . Note that the encoder has access to the output of the feedback link with one time-step delay. At time  $n$ , the decoder takes all the channel outputs  $Y_1, Y_2, \dots, Y_n$  and then produces the decoded message  $\hat{M}$ . We present the time ordering of these random variables below.

$$M, X_1, Y_1, Z_1, X_2, Y_2, Z_2, \dots, X_{n-1}, Y_{n-1}, Z_{n-1}, X_n, Y_n, \hat{M} \quad (3.1)$$

Note that all initial conditions (e.g. channel, feedback link, channel input, etc.) are assumed to be known in prior by both the encoder and the decoder. Before entering the more technical part of this chapter, it is necessary to give a specific definition of “noisy feedback”.

**Definition 10** (*Noisy Feedback*) *The feedback link is noisy if for some time instant  $i$  there exists no deterministic function  $g_i$  such that*

$$g_i(X^i, Z^i, M) = Y^i. \quad (3.2)$$

*The feedback link is noiseless if it is not noisy.*

**Remark 11** *This definition states that, for noisy feedback links, not all the channel outputs can be exactly recovered at the encoder side and, therefore, the encoder and decoder lose mutual understanding. In other words, at time instant  $i + 1$ , the encoder cannot access to the past channel outputs  $Y^i$  through information  $(X^i, Z^i, M)$  to produce channel input  $X_{i+1}$ . We refer “perfect (ideal) feedback” to be the case of  $Z^i = Y^i$  for all time instant  $i$ . Essentially, noiseless*

feedback is equivalent to perfect feedback since, in both cases, the encoder can access to the channel outputs without any error.

**Example 12** Consider the feedback link as  $Z_i = Y_i + V_i$  where  $V_i$  denotes additive noise at time instant  $i$ . If channel outputs  $Y_i$  only takes value in a set of integers (i.e.  $\pm 1, \pm 2, \dots$ ) and  $V_i$  only takes value in  $\{\pm 0.2, \pm 0.4\}$ , then obviously the channel outputs can be exactly recovered at the encoder side. Thus, this feedback link is noiseless even though it is imperfect. In this thesis, we may use “perfect feedback” and “noiseless feedback” alternatively without affecting any result.

Next, we give a definition of *typical noisy feedback link* which will be studied in the rest of the thesis.

**Definition 13** (*Typical Noisy Feedback Link*) Given channel  $\{p(y_i|x^i, y^{i-1})\}_{i=1}^\infty$ , the noisy feedback link  $\{p(z_i|y^i, z^{i-1})\}_{i=1}^\infty$  is typical if it satisfies

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(Z^{i-1}|Y^{i-1}) > 0 \quad (3.3)$$

for any channel input distribution  $\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^\infty$ . The noisy feedback link is non-typical if it is not typical.

**Remark 14** This definition implies that the noise in the feedback link must be active consistently over time (e.g. not physically vanishing). In practice, the typical noisy feedback link is the most interesting case for study.

**Example 15** Consider a binary symmetric feedback link modeled as  $Z_i = Y_i \oplus V_i$  where noise  $V_i$  is i.i.d and takes value from  $\{0, 1\}$  with equal probability. Then we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(Z^{i-1}|Y^{i-1}) &= \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(V^{i-1}|Y^{i-1}) \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(V_{i-1}|Y^{i-1}) \\ &\stackrel{(a)}{=} \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(V_{i-1}) \\ &= 1 \end{aligned}$$

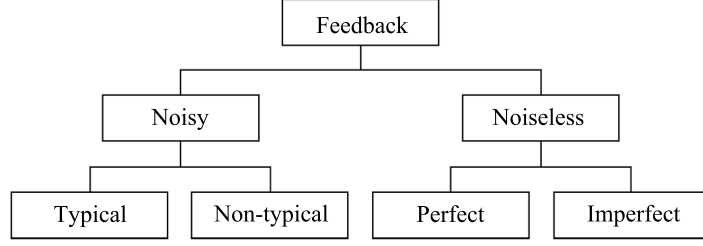


Figure 3.2 Family of feedback links in communication systems: the “typical noisy feedback” is the case which we are interested in.

where (a) follows from the fact that  $Y^{i-1}$  is independent from  $V_{i-1}$  due to one step delay. Therefore, this noisy feedback link is typical.

We summarize the family of the feedback link in Fig.3.2. In the sequel, the term “noisy feedback” refers to “typical noisy feedback” unless specified otherwise.

When there is no feedback from the channel output to the encoder, the maximum of mutual information (i.e.  $\lim_{n \rightarrow \infty} \max_{p(x^n)} I(X^n; Y^n)$ ) characterizes the maximum effective information flow through the channel with arbitrarily small probability of decoding error. This quantity is proved to be the capacity of the channel. When there is a noiseless feedback, supremizing directed information  $I(X^n \rightarrow Y^n)$  over  $p(x^n|y^n)$  with  $n \rightarrow \infty$  gives us the feedback capacity, e.g. Tatikonda and Mitter (2009), Kim (2008), Permuter et al. (2009). When there exists a noisy feedback, the appropriate measure/characterization of the effective information flow through the channel has been unknown until now. In the next section, we provide the missing measure.

We end this section with introducing the channel causality. We will say communication channel is causal if, for source input information  $\phi(n)$  (e.g. feedback information, message index, etc.), each channel input sequence  $x^n$  ( $x^n$  is a deterministic function of  $\phi(n)$ ) and the corresponding output sequence  $y^n$ ,

$$p(y_n | x^n(\phi(n)), y^{n-1}, \phi(n)) = p(y_n | x^n, y^{n-1}).$$

The idea of this definition is that the source information  $\phi(n)$  should be thought of as generated prior to the production of the channel output  $y_n$  and the channel should be aware of such

information only via its past channel inputs and outputs and its current input. Now we assume the channel and feedback link modeled in Fig. 3.1 are causal. It is then straightforward to have the following facts,

$$\begin{aligned} p(Y_n|X^n, Y^{n-1}, Z^{n-1}, M) &= p(Y_n|X^n, Y^{n-1}), \\ p(Z_n|Y^n, Z^{n-1}, X^n, M) &= p(Z_n|Y^n, Z^{n-1}). \end{aligned}$$

### 3.2 Residual Directed Information

Now, we propose a new information theoretic concept to capture the effective information flow (i.e.  $I(M; Y^n)$ ) in noisy feedback channels. Consider random variables  $(X^n, Y^n, W^n)$  with time order

$$W_1, X_1, Y_1, W_2, X_2, Y_2, \dots, W_{n-1}, X_{n-1}, Y_{n-1}, W_n, X_n, Y_n. \quad (3.4)$$

Based on the concepts of “directed information” and the “causal conditioning directed information”, the *residual directed information* and its density from  $X^n$  to  $Y^n$  w.r.t.  $W^n$  is defined as follows.

**Definition 16** (*Residual Directed Information and Its Density*) Consider the random variables  $(X^n, Y^n, W^n)$  with time order (3.4). The residual directed information from  $X^n$  to  $Y^n$  w.r.t.  $W^n$  is defined by

$$I^R(X^n(W^n) \rightarrow Y^n) = I(X^n \rightarrow Y^n) - I(X^n \rightarrow Y^n || W^n); \quad (3.5)$$

the corresponding residual directed information density is defined by

$$i^R(X^n(W^n) \rightarrow Y^n) = i(X^n \rightarrow Y^n) - i(X^n \rightarrow Y^n || W^n)$$

By applying this new concept to our specific noisy feedback channels (Fig. 3.1) with time ordering of random variables in (3.1), we have

$$I^R(X^n(M) \rightarrow Y^n) = I(X^n \rightarrow Y^n) - I(X^n \rightarrow Y^n || M). \quad (3.6)$$

As the message index  $M$  is given in prior and not evolving/changing with time, there exists no causality issue when we take the conditioning. Thus, equivalently, we have

$$I^R(X^n(M) \rightarrow Y^n) = I(X^n \rightarrow Y^n) - I(X^n \rightarrow Y^n | M). \quad (3.7)$$

The following theorem shows that the residual directed information,  $I^R(X^n(M) \rightarrow Y^n)$ , captures the mutual information between the message and the channel outputs which we refer to be the *effective information flow*.

**Theorem 17** *If  $X^n$  and  $Y^n$  are the inputs and outputs, respectively, of a discrete channel with noisy feedback, as shown in Fig.3.1, then*

$$I(M; Y^n) = I^R(X^n(M) \rightarrow Y^n) = I(X^n \rightarrow Y^n) - I(X^n \rightarrow Y^n | M).$$

**Proof.**

$$\begin{aligned}
& I(M; Y^n) \\
&= H(Y^n) - H(Y^n | M) \\
&= \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M) \\
&= \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M, X^i) - \left( \sum_{i=1}^n H(Y_i | Y^{i-1}, M) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M, X^i) \right) \\
&\stackrel{(a)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Y_i | Y^{i-1}, X^i) - \left( \sum_{i=1}^n H(Y_i | Y^{i-1}, M) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M, X^i) \right) \\
&= \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}) - \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}, M) \\
&= I(X^n \rightarrow Y^n) - I(X^n \rightarrow Y^n | M) \\
&\stackrel{(b)}{=} I^R(X^n(M) \rightarrow Y^n)
\end{aligned}$$

where (a) follows from the causality of the forward channel, i.e., the Markov chain  $M - (X^i, Y^{i-1}) - Y_i$ . Line (b) follows from the definition of the residual directed information. ■

**Remark 18** *This theorem implies that, for noisy feedback channels, the directed information  $I(X^n \rightarrow Y^n)$  captures both the effective information flow (i.e.  $I(M; Y^n)$ ) generated by the message and the redundant information flow (i.e.  $I(X^n \rightarrow Y^n | M)$ ) generated by the feedback noise (dummy message). Since only  $I(M; Y^n)$  is the relevant quantity for channel capacity, the well-known directed information clearly fails to characterize the noisy feedback capacity.*

In the following corollary, we explore some properties of the residual directed information  $I^R(X^n(M) \rightarrow Y^n)$ .

**Corollary 19** *The residual directed information  $I^R(X^n(M) \rightarrow Y^n)$  satisfies the following properties:*

1.  $I^R(X^n(M) \rightarrow Y^n) \geq 0$  (with equality if and only if the message set  $M$  and channel outputs  $Y^n$  are independent.)
2.  $I^R(X^n(M) \rightarrow Y^n) \leq I(X^n \rightarrow Y^n) \leq I(X^n; Y^n)$ .

*The first equality holds if the feedback is perfect. The second equality holds if there is no feedback.*

**Proof.** 1). Following from Theorem 17,  $I^R(X^n(M) \rightarrow Y^n) = I(M; Y^n) \geq 0$ . The necessary and sufficient condition of  $I^R(X^n(M) \rightarrow Y^n) = 0$  is obvious by looking at  $I(M; Y^n)$ .

2). Since  $I(X^n \rightarrow Y^n | M) = \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}, M) \geq 0$  (equality holds for the perfect feedback case),

$$I^R(X^n(M) \rightarrow Y^n) = I(X^n \rightarrow Y^n) - I(X^n \rightarrow Y^n | M) \leq I(X^n \rightarrow Y^n)$$

The proof of the second inequality  $I(X^n \rightarrow Y^n) \leq I(X^n; Y^n)$  is presented in Massey (1990). ■

Now, we give bounds on the first and second moments of the density function.

**Proposition 20** *The following inequality holds for any channel input distribution, channels and feedback links.*

- 1)  $\mathbb{E}[i^R(X^n(M) \rightarrow Y^n)] \leq \log |\mathcal{Y}^n|$
- 2)  $\text{Var}[i^R(X^n(M) \rightarrow Y^n)] \leq 2|\mathcal{Y}^n|$

**Proof.** See Appendix. ■

### 3.3 Information Flow Decomposition in Noisy Feedback Channels

Having proposed a new concept, residual directed information, we find out that the directed information is not the right quantity to work on in order to characterize the capacity of noisy feedback channels. In particular, in noiseless feedback setting, we have  $I(M; Y^n) = I(X^n \rightarrow Y^n)$ , implying that the directed information flow from  $X^n$  to  $Y^n$  is fully used for delivering the message. However, as it is shown in Theorem 17, this fact does not hold due to the extra term  $I(X^n \rightarrow Y^n | M)$ . What does this term imply? or what are the redundant information flows captured by this extra term? We give an answer to this question in this section and, as a consequence, we have a clear picture of the information flow in noisy feedback channels.

**Theorem 21** (*Information Flow Decomposition Equality(IFDE)*) *If  $X^n$  and  $Y^n$  are the inputs and outputs, respectively, of a discrete channel with noisy feedback, and  $Z^n$  is the outputs of the feedback link, as shown in Fig.3.1, then*

$$I(X^n \rightarrow Y^n) = I(M; Y^n) + I(Z^{n-1} \rightarrow Y^n) + I(Z^{n-1}; M | Y^n).$$

**Proof.** Based on Theorem 17, we only need to show

$$I(X^n \rightarrow Y^n | M) = I(Z^{n-1} \rightarrow Y^n) + I(Z^{n-1}; M | Y^n).$$

We begin with  $I(Z^{n-1}; M|Y^n)$ , that is,

$$\begin{aligned}
& I(Z^{n-1}; M|Y^n) \\
&= I(Z^{n-1}, Y^n; M) - I(Y^n; M) \\
&= \sum_{i=1}^n I(Z_{i-1}, Y_i; M|Z^{i-2}, Y^{i-1}) - I(Y^n; M) \\
&= \sum_{i=1}^n H(Z_{i-1}, Y_i|Z^{i-2}, Y^{i-1}) - H(Z_{i-1}, Y_i|Z^{i-2}, Y^{i-1}, M) - I(Y^n; M) \\
&= \sum_{i=1}^n H(Y_i|Z^{i-1}, Y^{i-1}) + H(Z_{i-1}|Z^{i-2}, Y^{i-1}) - H(Y_i|Z^{i-1}, Y^{i-1}, M) \\
&\quad - H(Z_{i-1}|Z^{i-2}, Y^{i-1}, M) - I(Y^n; M) \\
&\stackrel{(a)}{=} \sum_{i=1}^n H(Y_i|Z^{i-1}, Y^{i-1}) + H(Z_{i-1}|Z^{i-2}, Y^{i-1}) - H(Y_i|Z^{i-1}, Y^{i-1}, M) \\
&\quad - H(Z_{i-1}|Z^{i-2}, Y^{i-1}) - I(Y^n; M) \\
&= \sum_{i=1}^n H(Y_i|Z^{i-1}, Y^{i-1}) - H(Y_i|Z^{i-1}, Y^{i-1}, M) - I(Y^n; M) \\
&= \sum_{i=1}^n H(Y_i|Z^{i-1}, Y^{i-1}) - H(Y_i|X^i(Z^{i-1}, M), Z^{i-1}, Y^{i-1}, M) - I(Y^n; M) \\
&\stackrel{(b)}{=} \sum_{i=1}^n H(Y_i|Z^{i-1}, Y^{i-1}) - H(Y_i|X^i, Y^{i-1}, M) - H(Y^n) + H(Y^n|M)
\end{aligned}$$

where line (a) follows from the causality of the feedback link, i.e., the Markov chain  $M - (Z^{i-1}, Y^i) - Z_i$ . Line (b) follows from the causality of the forward channel, i.e., the Markov chain  $(M, Z^{i-1}) - (X^i, Y^i) - Y_i$ .

Next, we have

$$\begin{aligned}
& I(Z^{n-1} \rightarrow Y^n) \\
&= H(Y^n) - H(Y^n||Z^{n-1}) \\
&= H(Y^n) - \sum_{i=1}^n H(Y_i|Z^{i-1}, Y^{i-1})
\end{aligned}$$



Then the sum

$$\begin{aligned}
& I(Z^{n-1} \rightarrow Y^n) + I(Z^{n-1}; M|Y^n) \\
&= \sum_{i=1}^n H(Y_i|Y^{i-1}, M) - H(Y_i|X^i, Y^{i-1}, M) \\
&= \sum_{i=1}^n I(X^i, Y_i|Y^{i-1}, M) \\
&= I(X^n \rightarrow Y^n|M)
\end{aligned}$$

The proof is complete. ■

It is clear to check that, if the feedback channel is perfect (i.e.  $Z_i = Y_i$  for all  $i$ ), the last two information flows (shown in Theorem 21) delivered in the forward channel turn out to be zero. As a result, the quantity  $I(Z^{n-1} \rightarrow Y^n)$  can be treated as a measure of the amount of information delivered in the forward channel as a result of adding uncertainties in the feedback channel. In addition, the quantity  $I(Z^{n-1}; M|Y^n)$  can be treated as a measure of the amount of the interference between the uncertainties in the feedback channel and the message while both of them are delivered in the forward channel.

To gain more insight in the information flow pattern, we next investigate channels with additive noise feedback and analyze its information flow on a dependency graph. See Fig.3.3. We present the time ordering of these random variables below.  $Z_i$  is not shown in the time ordering since we have  $Z_i = Y_i + V_i$ .

$$M, X_1, Y_1, V_1, X_2, Y_2, V_2, \dots, X_{n-1}, Y_{n-1}, V_{n-1}, X_n, Y_n, \hat{M}$$

**Corollary 22** *If  $X^n$  and  $Y^n$  are the inputs and outputs, respectively, of a discrete channel with additive noise feedback, as shown in Fig.3.3, then*

$$I(X^n \rightarrow Y^n) = I(M; Y^n) + I(V^{n-1}; Y^n) + I(M; V^{n-1}|Y^n)$$

**Proof.** See Appendix. ■

Corollary 22 allows us to explicitly interpret the information flow on a dependency graph (e.g.  $N = 3$ ). See Fig.3.4. The solid lines from message  $M$  to sequence  $X^3$  represent the

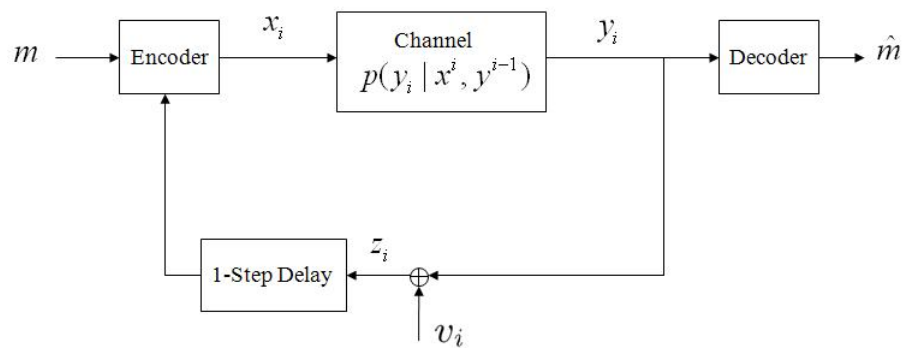


Figure 3.3 Communication Channels with additive noise feedback

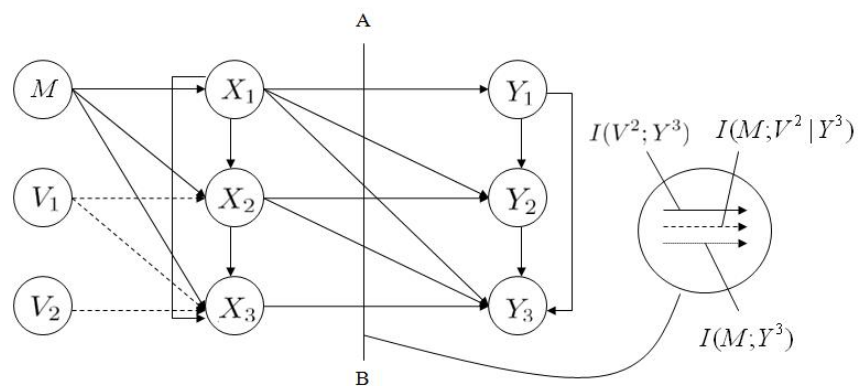


Figure 3.4 The information flow of channels with additive noise feedback

dependence of  $X^3$  on  $M$ . The dotted lines from additive noise  $V^2$  to sequence  $X^3$  represent the dependence of  $X^3$  on  $V^2$ . The dependence of the channel inputs  $X^3$  on the channel outputs  $Y^2$  is not shown in the graph since the directed information only captures the information flow from  $X^3$  to  $Y^3$  (Massey (1990)). As it is shown in the zoomed circle, the directed information flow from  $X^3$  to  $Y^3$  (through cut  $A - B$ ) implicitly contains three sub-information flows wherein the mutual information  $I(M; Y^3)$  and  $I(V^2; Y^3)$  measure the message-delivery and the noise-delivery information flows, respectively. The feedback noise  $V^2$  is treated as a dummy message which also needs to be recovered by the decoder. The conditional mutual information  $I(M; V^2 | Y^3)$  quantifies the mixed information flow between the message-transmitting and noise-transmitting flows. Essentially, the second term in the residual directed information (i.e.  $I(X^n \rightarrow Y^n | M)$ ) precisely captures the non-message transmitting information flows (i.e.  $I(V^{n-1}; Y^n)$  and  $I(M; V^{n-1} | Y^n)$ ). Therefore, the residual directed information should be a proper measure to work with for channels with noisy feedback.

### 3.4 Conclusions

In this chapter, we have proposed a new concept, *residual directed information*, to capture the message-delivery quantity in channels with noisy feedback. The new concept indicates in principle the failure of using the well-know mutual information or directed information to characterize noisy feedback capacity. Motivated by this new concept, we next derived the main result, information flow decomposition equality, which reveals the information flow pattern in noisy feedback channels.

In sum, understanding the information flow in noisy feedback channels leads us to a higher level to investigate the noisy feedback problem and performs as the basis to develop fruitful results (to be seen later in this thesis).

## CHAPTER 4. A CHANNEL CODING THEOREM AND BOUNDS ON THE CAPACITY OF FINITE-ALPHABET CHANNELS WITH NOISY FEEDBACK

### 4.1 Introduction

In this section, we first derive a channel coding theorem for finite-alphabet channels with noisy feedback Li and Elia (2011c). The result is of theoretical value as it fits in the general theory already developed for non-feedback channels. However the characterization of the capacity has infinite multi-letter form and is not computable in general. But as a mediate technical step, this channel coding theorem allows us to derive bounds in terms of the well-known causal conditioning directed information.

To be concrete, for general non-feedback channels, Verdú and Han have characterized the general channel capacity by invoking Feinstein's lemma. Tatikonda and Mitter (2009) extended this approach to finite-alphabet channels with noiseless feedback and characterized the capacity in term of the directed information. The main idea therein is to convert the channel coding problem with noiseless feedback into an equivalent channel coding problem without feedback by considering code-functions instead of code-words. In fact, code-functions can be treated as a generalization of code-words. In this chapter, we extend this idea to noisy feedback settings to obtain a channel coding theorem.

As it is shown in Chapter 3, for channels with noisy feedback, the directed information is not a proper quantity to characterize the channel capacity. We then proposed a new concept, residual directed information, and showed that this quantity equals the mutual information between the message and the channel outputs (i.e. the information received by the decoder). As the first and important application of this new concept, in this chapter, we show that the

residual directed information can be used to characterize the capacity of finite alphabet channels with noisy feedback, by invoking the code-function approach used in Tatikonda and Mitter (2009). As it will be shown, this characterization has nice features and provides much insight in the noisy feedback capacity. We then propose capacity bounds which are characterized by the causal conditioning directed information. Finally, we investigate the special class of channels, finite-state finite-alphabet channels, and provide an upper bound in form of a finite dimensional optimization on its capacity.

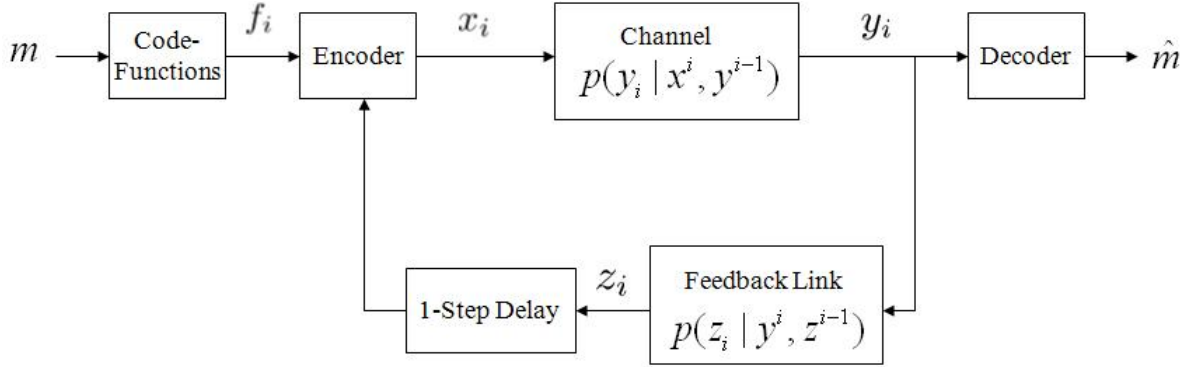


Figure 4.1 Channels with noisy feedback (a code-function representation)

## 4.2 Problem Formulation and Preliminaries

We first formulate the channel coding problem. Here, we require the use of code-functions as opposed to codewords, as shown in Fig.4.1. Briefly, at time 0, we choose a message from a message set  $M$ . This message is associated with a sequence of code-functions. Then from time 1 to  $n$ , we use the channels to transmit information sequentially based on the corresponding code-function. At time  $n + 1$ , we decode the message as  $\hat{M}$ . We now give a formal definition of this communication scheme, which extends the description presented in Tatikonda and Mitter (2009).

**Definition 23** (*Communication Scheme for Channels with Noisy Feedback: A Code-function*

*Representation)*

1. A message index  $m \in \{1, 2, \dots, M\}$
2. A channel code-function is a sequence of  $n$  deterministic measurable maps  $f^n = \{f_i\}_{i=1}^n$  ( $f \in \mathcal{F}$ ) such that  $f_i : \mathcal{Z}^{i-1} \rightarrow \mathcal{X}$  which takes  $z^{i-1} \mapsto x_i$ .
3. A channel encoder is a set of  $M$  channel code-functions, denoted by  $\{f^n[m]\}_{m=1}^M$ .
4. A channel is a family of conditional probability  $\{p(y_i|x^i, y^{i-1})\}_{i=1}^n$ .
5. A noisy feedback link is a family of conditional probability  $\{p(z_i|y^i, z^{i-1})\}_{i=1}^n$ .
6. A channel decoder is a map  $g$  which takes  $y^n \mapsto m$ .

Based on the above communication scheme, we redefine the channel code and  $\epsilon$ -achievable rate in terms of code-functions.

**Definition 24** (*Channel Code*) A  $(n, M, \epsilon)$  channel code over time horizon  $n$  consists of  $M$  code-functions  $\{f^n[m]\}_{m=1}^M$ , a channel decoder  $g$ , and an error probability satisfying

$$\frac{1}{M} \sum_{m=1}^M p(m \neq g(y^n)|m) \leq \epsilon$$

**Definition 25** ( $\epsilon$ -achievable Rate)  $R \geq 0$  is an  $\epsilon$ -achievable rate if, for every  $\epsilon > 0$ , there exist, for all sufficiently large  $n$ , a  $(n, M, \epsilon)$  channel code with rate

$$\frac{\log M}{n} \geq R - \epsilon$$

The maximum  $\epsilon$ -achievable rate is called the  $\epsilon$ -capacity, denoted by  $C_{fb}^{noisy}(\epsilon)$ . The channel capacity  $C_{fb}^{noisy}$  is defined as the maximal rate that is  $\epsilon$ -achievable for all  $0 < \epsilon < 1$ . Clearly,  $C_{fb}^{noisy} = \lim_{\epsilon \rightarrow 0} C_{fb}^{noisy}(\epsilon)$

The channel coding problem is to search for a sequence of  $(n, M, \epsilon)$  channel codes under which the achievable rate is maximized as  $n \rightarrow \infty$ . In order to construct a general channel coding theorem (i.e. no restrictions on channels and input/output alphabets, such as stationary, ergodic,  $\dots$ ), we introduce the following two probabilistic limit operations (Han (2003)).

**Definition 26** (*Probabilistic Limit*) The limit superior in probability for any sequence  $(X_1, X_2, \dots)$  is defined by

$$p - \limsup_{n \rightarrow \infty} X_n = \inf\{\alpha \mid \lim_{n \rightarrow \infty} \text{Prob}\{X_n > \alpha\} = 0\}$$

Similarly, the limit inferior in probability for any sequence  $(X_1, X_2, \dots)$  is defined by

$$p - \liminf_{n \rightarrow \infty} X_n = \sup\{\beta \mid \lim_{n \rightarrow \infty} \text{Prob}\{X_n < \beta\} = 0\}$$

Next, we introduce some notations.

$$\begin{aligned} \underline{I}(X; Y) &= p - \liminf_{n \rightarrow \infty} \frac{1}{n} i(X^n; Y^n) \\ \bar{I}(X; Y) &= p - \limsup_{n \rightarrow \infty} \frac{1}{n} i(X^n; Y^n) \\ \underline{I}^R(X(F) \rightarrow Y) &= p - \liminf_{n \rightarrow \infty} \frac{1}{n} i^R(X^n(F^n) \rightarrow Y^n) \\ \bar{I}^R(X(F) \rightarrow Y) &= p - \limsup_{n \rightarrow \infty} \frac{1}{n} i^R(X^n(F^n) \rightarrow Y^n) \end{aligned}$$

Following the idea in Tatikonda and Mitter (2009), it is convenient to consider the noisy feedback channel problem as a regular nonfeedback problem from the input alphabet  $\mathcal{F}$  to the output alphabet  $\mathcal{Y}$  as shown in Fig.4.1. This consideration provides us with an approach to prove the channel coding theorem for channels with noisy feedback. Recall that the capacity of nonfeedback channels is characterized as follows (Verdú and Han (1994)).

**Theorem 27** (*Non-feedback Channel Capacity*) *For any channel with arbitrary input and output alphabets  $\mathcal{F}$  and  $\mathcal{Y}$ , the channel capacity  $C$  is given by*

$$C = \sup_F \underline{I}(F; Y)$$

where  $\sup_F$  denotes the supremum with respect to all the input processes  $F$ .

However, before applying the above result, we need to understand the inherent connection between the equivalent nonfeedback channel and the original channel with noisy feedback link. Moreover, as supremizing the mutual information over code-function  $F$  is inconvenient, we need create a connection between the nonfeedback channel input distribution  $\{p(f^n)\}$  and the original channel input distribution such that we can still work on the original channel input. These two issues are the main technical steps toward the channel coding theorem. We provide these results as lemmas in the next subsection. Then, we prove the channel coding theorem along the lines of the proof of Theorem 27.

### 4.2.1 Technical Lemmas

We first show an equality of information densities between the nonfeedback channel  $\mathcal{F}^n \rightarrow \mathcal{Y}^n$  and the original channel  $\mathcal{X}^n \rightarrow \mathcal{Y}^n$ .

**Lemma 28**

$$i(F^n; Y^n) = i^R(X^n(F^n) \rightarrow Y^n)$$

where  $i^R(X^n(F^n) \rightarrow Y^n)$  is defined as

$$i^R(X^n(F^n) \rightarrow Y^n) = i(X^n \rightarrow Y^n) - i(X^n \rightarrow Y^n || F^n).$$

**Proof.**

$$\begin{aligned}
i(F^n; Y^n) &= \log \frac{p(F^n, Y^n)}{p(F^n)p(Y^n)} \\
&= \log \frac{\prod_{i=1}^n p(F_i, Y_i | F^{i-1}, Y^{i-1})}{p(F^n)p(Y^n)} \\
&= \log \frac{\prod_{i=1}^n p(Y_i | F^i, Y^{i-1}) p(F_i | F^{i-1}, Y^{i-1})}{p(F^n)p(Y^n)} \\
&\stackrel{(a)}{=} \log \frac{\prod_{i=1}^n p(Y_i | F^i, Y^{i-1}) p(F_i | F^{i-1})}{p(F^n)p(Y^n)} \\
&= \log \frac{\tilde{p}(Y^n | F^n, X^n)}{p(Y^n)} - \log \frac{\tilde{p}(Y^n | F^n, X^n)}{\prod_{i=1}^n p(Y_i | F^i, Y^{i-1})} \\
&= \log \frac{\prod_{i=1}^n p(Y_i | F^i, X^i, Y^{i-1})}{p(Y^n)} - \log \frac{\tilde{p}(Y^n | F^n, X^n)}{\prod_{i=1}^n p(Y_i | Y^{i-1}, F^i)} \\
&\stackrel{(b)}{=} \log \frac{\prod_{i=1}^n p(Y_i | X^i, Y^{i-1})}{p(Y^n)} - \log \frac{\tilde{p}(Y^n | F^n, X^n)}{\tilde{p}(Y^n | F^n)} \\
&= \log \frac{\tilde{p}(Y^n | X^n)}{p(Y^n)} - \log \frac{\tilde{p}(Y^n | F^n, X^n)}{\tilde{p}(Y^n | F^n)} \\
&= i(X^n \rightarrow Y^n) - i(X^n \rightarrow Y^n || F^n) \\
&= i^R(X^n(F^n) \rightarrow Y^n)
\end{aligned}$$

where (a) follows from the fact that no feedback exists from  $\mathcal{Y}$  to  $\mathcal{F}$ . Line (b) follows from the Markov chain  $F^i - (X^i, Y^{i-1}) - Y_i$ . ■

In the next lemma, we show that there exists a suitable construction of  $p(f^n)$  such that the induced channel input distribution equals the original channel input distribution. As we will



see, this result allows us to work on the channel input distributions instead of code-function distributions.

**Lemma 29** *Given a channel  $\{p(y_i|x^i, y^{i-1})\}_{i=1}^n$ , a feedback link  $\{p(z_i|y^i, z^{i-1})\}_{i=1}^n$ , a channel input distribution  $\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^n$  and a sequence of code-function distributions  $\{p(f_i|f^{i-1})\}_{i=1}^n$ , the induced channel input distribution  $\{p_{ind}(x_i|x^{i-1}, z^{i-1})\}_{i=1}^n$  (induced by  $\{p(f_i|f^{i-1})\}_{i=1}^n$ ) equals the original channel input distribution  $\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^n$  if and only if the sequence of code-function distributions  $\{p(f_i|f^{i-1})\}_{i=1}^n$  is good with respect to  $\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^n$ . One choice of such a sequence of code-function distributions is as follows,*

$$p(f_i|f^{i-1}) = \prod_{z^{i-1}} p(f_i(z^{i-1})|f^{i-1}(z^{i-2}), z^{i-1}). \quad (4.1)$$

We refer the readers to Definition 5.1, Lemma 5.1 and 5.4 in Tatikonda and Mitter (2009) for the concept “good with respect to” and the proof of the above lemma. According to Lemma 29, it is straightforward to obtain the following result which plays an essential role in the channel coding theorem.

**Lemma 30** *For channels with noisy feedback,*

$$\begin{aligned} & p(x^n, y^n, f^n) \\ &= \prod_{i=1}^n \prod_{z^{i-1}} \underbrace{p(f_i(z^{i-1})|f^{i-1}(z^{i-2}), z^{i-1})}_{\text{Encoding}} \sum_{z^n \in \{\mathcal{Z}^n : x^n = f^n(z^{n-1})\}} \prod_{i=1}^n \underbrace{p(z_i|y^i, z^{i-1})}_{\text{Feedback link}} \underbrace{p(y_i|f^i(z^{i-1}), y^{i-1})}_{\text{Channel}} \end{aligned}$$

The proof is shown in the Appendix. This lemma implies that  $\underline{I}^R(X(F) \rightarrow Y)$  only depends on channel input distribution  $\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^\infty$ .

### 4.3 Channel Coding Theorem

Now we show a general channel coding theorem in terms of the residual directed information.

**Theorem 31** (*Channel Coding Theorem*) For channels with noisy feedback,

$$C_{fb}^{noisy} = \sup_X \underline{I}^R(X(F) \rightarrow Y) \quad (4.2)$$

where  $\sup_X$  means that supremum is taken over all possible channel input distributions  $\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^\infty$ .

The proof comes along the proof of Theorem 27 in Verdú and Han (1994) and hence is presented in the Appendix. Theorem 31 indicates that, besides capturing the effective information flow of channels with noisy feedback, the residual directed information is also beneficial for characterizing the capacity. Although formula (4.2) may not be the only or the simplest characterization of the noisy feedback capacity, it provides benefits in many aspects. We herein present two of them as follows.

1. Measurements of information flows: Let  $p^*$  be the optimal solution of formula (4.2). Then we obtain that, when the channel is used at capacity, the total transmission rate in the forward channel is in fact  $\underline{I}(X \rightarrow Y)|_{p^*}$ <sup>1</sup> instead of  $C_{fb}^{noisy}$  and the difference between them (i.e. redundant transmission rate) is  $\underline{I}(X \rightarrow Y|F)|_{p^*}$ . These numerical knowledge might be crucial in system design and evaluation.
2. Induced capacity bounds: Let  $q^* = \arg \sup_X \underline{I}(X \rightarrow Y)$  where supremum is taken over all possible channel input distributions  $\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^\infty$ . Since code-function  $F$  is not involved at this point, the computation complexity is significantly reduced. Based on Theorem 31, it is straightforward to obtain  $\underline{I}(X \rightarrow Y)|_{q^*}$  and  $\underline{I}^R(X(F) \rightarrow Y)|_{q^*}$  as upper<sup>2</sup> and lower bounds on the capacity, respectively. Further, the gap between the bounds is  $\underline{I}(X \rightarrow Y|F)|_{q^*}$ , which is definitely a tightness evaluation of the bounds.

#### 4.4 Capacity Bounds

As it is shown, the capacity characterization in Theorem 31 is not computable in general due to the probabilistic limit and code-functions. This motivates us to explore some conditions

<sup>1</sup> $\underline{I}(X \rightarrow Y)|_{p^*}$  denotes that the value is evaluated at channel input distributions  $p^*$ .

<sup>2</sup>Note that  $\underline{I}(X \rightarrow Y)|_{q^*} = \sup_{\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^\infty} \underline{I}(X \rightarrow Y) \leq C_{FB} = \sup_{\{p(x_i|x^{i-1}, y^{i-1})\}_{i=1}^\infty} \underline{I}(X \rightarrow Y)$  where  $C_{FB}$  is the corresponding perfect feedback capacity. Therefore this upper bound is in general better than  $C_{FB}$ .

under which the previous characterization can be simplified or to look at some computable bounds instead. Toward this end, we first introduce a strong converse theorem under which the “probabilistic limit” can be replaced by the “normal limit”. We then turn to characterize a pair of upper and lower bounds which is much easier to compute and tight in certain practical situations.

**Definition 32** (*Strong Converse*) A channel with noisy feedback capacity  $C_{fb}^{noisy}$  has a strong converse if for any  $R > C_{fb}^{noisy}$ , every sequence of channel codes  $\{(n, M_n, \epsilon_n)\}_{n=1}^{\infty}$  with

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M_n \geq R$$

satisfies  $\lim_{n \rightarrow \infty} \epsilon_n = 1$

**Theorem 33** (*Strong Converse Theorem*) A channel with noisy feedback capacity  $C_{fb}^{noisy}$  satisfies the strong converse property if and only if

$$\sup_X \underline{I}^R(X(F) \rightarrow Y) = \sup_X \bar{I}^R(X(F) \rightarrow Y)^3 \quad (4.3)$$

Furthermore, if the strong converse property holds, we have

$$C_{fb}^{noisy} = \sup_X \lim_{n \rightarrow \infty} \frac{1}{n} I^R(X^n(F^n) \rightarrow Y^n).$$

The proof directly follows from chapter 3.5 in Han (2003) by appropriate replacement of  $i^R(X^n(F^n) \rightarrow Y^n)$  on  $i(F^n; Y^n)$ . This theorem gives us an important message that, for channels satisfying the strong converse property, we may compute the noisy feedback capacity by taking the normal limit instead of the probabilistic limit. How to further simplify the capacity characterization will be explored in the future.

We next propose an upper bound on the noisy feedback capacity.

**Theorem 34** (*Upper Bound*)<sup>4</sup>

$$\bar{C}_{FB}^{noise} = \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n || Z^{n-1}) \quad (4.4)$$

<sup>3</sup>This condition can be alternatively expressed as  $\sup_X \underline{I}(F; Y) = \sup_X \bar{I}(F; Y)$ . Since the computation complexity difference between the mutual information and residual directed information is not justified, either condition is a candidate for check. Note that how to check the strong converse is out of the scope of this chapter.

<sup>4</sup>As we will see from the proof, this upper bound holds for any finite-alphabet channel with or without strong converse property.

where  $\bar{C}_{FB}^{noise}$  denotes the upper bound of the capacity and the supremum is taken over all possible channel input distribution  $\{p(x_i|x^{i-1}, z^{i-1})\}_{i=1}^\infty$ .

Note that this upper bound holds for channels with strong converse or not. We need the following lemma before showing the proof of Theorem 34.

**Lemma 35**

$$I(F^n; Y^n) = I^R(X^n(F^n) \rightarrow Y^n) = I(X^n \rightarrow Y^n \| Z^{n-1}) - I(F^n; Z^{n-1} | Y^n)$$

**Proof.** See Appendix. ■

Now we present the proof of Theorem 34 as follows.

**Proof.** Recall Lemma A1 in Han and Verdú (1993), we have  $\underline{I}(F; Y) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} I(F^n; Y^n)$  for any sequence of joint probability. That is,  $\underline{I}^R(X(F) \rightarrow Y) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} I^R(X^n(F^n) \rightarrow Y^n)$ . Then by Lemma 35,

$$\begin{aligned} C_{fb}^{noisy} &\leq \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I^R(X^n(F^n) \rightarrow Y^n) \\ &= \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} (I(X^n \rightarrow Y^n \| Z^{n-1}) - I(F^n; Z^{n-1} | Y^n)) \\ &\leq \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n \| Z^{n-1}) \end{aligned} \tag{4.5}$$

■

**Corollary 36** Assume that there is an independent additive noise feedback (Fig.3.4), then

$$\bar{C}_{FB}^{noise} = \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n \| V^{n-1})$$

where  $\sup_X$  means that supremum is taken over all possible channel input distribution  $\{p(x_i|x^{i-1}, y^{i-1} + v^{i-1})\}_{i=1}^\infty$ .

**Proof.**

$$\begin{aligned} I(X^n \rightarrow Y^n \| Z^{n-1}) &= \sum_{i=1}^n I(X^i, Y_i | Y^{i-1}, Z^{i-1}) \\ &= \sum_{i=1}^n I(X^i, Y_i | Y^{i-1}, V^{i-1}) \\ &= I(X^n \rightarrow Y^n \| V^{n-1}) \end{aligned}$$

■

Next, we show a lower bound on the capacity for strong converse channels with additive noise feedback. Although proposing a particular coding scheme is a standard approach to obtain a lower bound on the capacity, it is not clearly doable for noisy feedback settings. We herein propose a lower bound from another route which is not restricted to any specific coding scheme. In addition, this lower bound has nice features and its own advantages.

**Theorem 37** (*Lower Bound*) *Assume that a channel with an independent additive noise feedback (Fig.3.3) satisfies the strong converse property. A lower bound on the noisy feedback capacity is given by*

$$\underline{C}_{FB}^{noise} = \bar{C}_{FB}^{noise} - \bar{h}(V)$$

where

$$\bar{h}(V) = \limsup_{n \rightarrow \infty} \frac{1}{n} H(V^{n-1}).$$

**Proof.** We need to show that, for any  $\delta > 0$ , there exists a sequence of  $(n, M, \epsilon_n)$  channel codes ( $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ ) with transmission rate

$$\begin{aligned} R &= \bar{C}_{FB}^{noise} - \bar{h}(V) - \delta \\ &= \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n || V^{n-1}) - \bar{h}(V) - \delta. \end{aligned}$$

Now, for any fixed  $\delta > 0$ , we take  $\xi$  satisfying  $0 < \xi < \delta$  and let  $X_\xi$  be a sequence of channel input distributions  $\{p(x_i | x^{i-1}, z^{i-1})\}_{i=1}^\infty$  satisfying

$$\left( \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n || Z^{n-1}) \right) \Big|_{X=X_\xi} = \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n || Z^{n-1}) - \xi \quad (4.6)$$

where  $(\liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n || Z^{n-1}))|_{X=X_\xi}$  denotes that  $\liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n || Z^{n-1})$  is evaluated at  $X = X_\xi$ . According to the definition of supremum, the existence of  $X_\xi$  is guaranteed. Since for strong converse channels we have

$$C_{fb}^{noisy} = \sup_X \lim_{n \rightarrow \infty} \frac{1}{n} I^R(X^n(F^n) \rightarrow Y^n),$$

we know that, for any  $\delta > 0$ , there exist a sequence of  $(n, M, \epsilon_n)$  channel codes ( $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ ) with transmission rate

$$R = \left( \lim_{n \rightarrow \infty} \frac{1}{n} I^R(X^n(F^n) \rightarrow Y^n) \right) \Big|_{X=X_\xi} - (\delta - \xi).$$

By Lemma 35,

$$\begin{aligned} R &= \left( \lim_{n \rightarrow \infty} \frac{1}{n} (I(X^n \rightarrow Y^n \| Z^{n-1}) - I(F^n; Z^{n-1} | Y^n)) \right) \Big|_{X=X_\xi} - (\delta - \xi) \\ &= \left( \lim_{n \rightarrow \infty} \frac{1}{n} (I(X^n \rightarrow Y^n \| Z^{n-1}) - H(Z^{n-1} | Y^n) + H(Z^{n-1} | Y^n, F^n)) \right) \Big|_{X=X_\xi} - (\delta - \xi) \\ &\geq \left( \liminf_{n \rightarrow \infty} \frac{1}{n} (I(X^n \rightarrow Y^n \| Z^{n-1}) - H(Z^{n-1} | Y^n)) \right) \Big|_{X=X_\xi} - (\delta - \xi) \\ &= \left( \liminf_{n \rightarrow \infty} \frac{1}{n} (I(X^n \rightarrow Y^n \| Z^{n-1}) - \sum_{i=1}^n H(Z_{i-1} | Z^{i-2}, Y^n)) \right) \Big|_{X=X_\xi} - (\delta - \xi) \\ &\geq \left( \liminf_{n \rightarrow \infty} \frac{1}{n} (I(X^n \rightarrow Y^n \| Z^{n-1}) - \sum_{i=1}^n H(Z_{i-1} | Z^{i-2}, Y^{i-1})) \right) \Big|_{X=X_\xi} - (\delta - \xi) \\ &\stackrel{(a)}{=} \left( \liminf_{n \rightarrow \infty} \frac{1}{n} (I(X^n \rightarrow Y^n \| Z^{n-1}) - \sum_{i=1}^n H(V_{i-1} | V^{i-2})) \right) \Big|_{X=X_\xi} - (\delta - \xi) \\ &\geq \left( \liminf_{n \rightarrow \infty} \frac{1}{n} (I(X^n \rightarrow Y^n \| Z^{n-1}) - H(V^{n-1})) \right) \Big|_{X=X_\xi} - (\delta - \xi) \\ &\geq \left( \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n \| Z^{n-1}) \right) \Big|_{X=X_\xi} + \liminf_{n \rightarrow \infty} -\frac{1}{n} H(V^n) - (\delta - \xi) \\ &= \left( \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n \| Z^{n-1}) \right) \Big|_{X=X_\xi} - \limsup_{n \rightarrow \infty} \frac{1}{n} H(V^n) - (\delta - \xi) \\ &\stackrel{(b)}{=} \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n \| Z^{n-1}) - \xi - \bar{h}(V) - (\delta - \xi) \\ &= \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n \| Z^{n-1}) - \bar{h}(V) - \delta \\ &\stackrel{(c)}{=} \sup_X \liminf_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n \| V^{n-1}) - \bar{h}(V) - \delta \end{aligned}$$

where (a) follows from the fact that  $Z_i = Y_i + V_i$  and the Markov Chain  $(Z^{i-1}, Y^i) - V^{i-1} - V_i$ .

Line (b) follows from equation (4.6). Line (c) follows from Corollary 3.

Since  $\delta$  can be arbitrarily small, the proof is complete. ■

**Remark 38** *This theorem reveals an important message that the gap between the proposed upper and lower bounds only depends on the feedback additive noise  $V$  (i.e. independent from*

the forward channel). Further, if the entropy rate of noise  $V$  goes to zero<sup>5</sup>, the proposed upper and lower bound converges and thus the capacity is known.

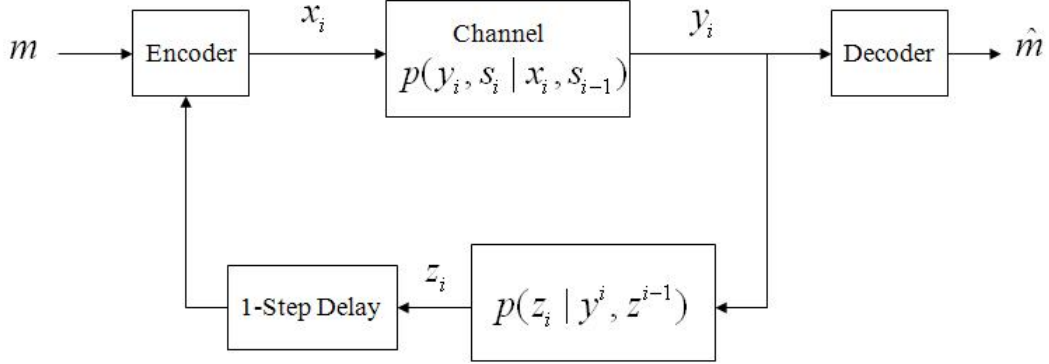


Figure 4.2 Finite state channels with noisy feedback

#### 4.5 Case Study: Capacity Bounds for Finite State Channels

In the previous section, we provide bounds, in terms of the causal conditioning directed information, on the noise feedback capacity. However, evaluate the limit value of these formulas is notoriously difficult in general. In this section, we consider a special class of noisy feedback channels, finite state channels (FSC), and then propose an upper bound on the capacity.

Firstly investigated in Gallager (1968), FSC are a class of channels rich enough to include channels with memory, e.g., channels with inter-symbol interference (ISI). FSC with feedback have attracted much attention in the recent decade. The capacity of some channels with channel state information at the receiver and transmitter was derived by Viswanathan (1999) and Caire and Shamai(Shitz) (1999). In Permuter et al. (2009), the capacity of FSC with deterministic feedback was characterized, which is a generalization of the non-feedback and noiseless feedback cases. In addition, Yang et al. (2005) computed the feedback FSC capacity under the assumption that the state channel is a deterministic function of the previous state

<sup>5</sup>In many practical situations, the entropy rate of the feedback noise is small. For example, if the feedback link only suffers intersymbol interference as illustrated in Chapter 4 Gallager (1968), the entropy rate turns out to be approximately 0.0808. Further, if the cardinality of  $V^\infty$  is finite (yet the feedback is still noisy), the entropy rate is clearly zero.

and input. Chen and Berger (2005) also computed the feedback FSC capacity but under the assumption that the state channel is a deterministic function of the output. In this section, we consider FSC with noisy feedback as shown in Fig 4.2. The input of the channel is denoted by  $\{X_i\}_{i=1}^\infty$ , and the output of the channel is denoted by  $\{Y_i\}_{i=1}^\infty$ . In addition, the channel states take values in a finite set of possible states  $\mathcal{S}$ . The channel is stationary and is characterized by a conditional probability assignment  $p(y_i, s_i|x_i, s_{i-1})$  that satisfies

$$p(y_i, s_i|x^i, s^{i-1}, y^{i-1}) = p(y_i, s_i|x_i, s_{i-1}),$$

where the initial state distribution is  $p(s_0)$ . An FSC is said to be without ISI if the input sequence does not affect the evolution of the state sequence, i.e.,  $p(s_i|s_{i-1}, x_i) = p(s_i|s_{i-1})$ .

The feedback channel is characterized in a general form  $p(z_i|y^i, z^{i-1})$  and is assumed to be stationary.

Define

$$\bar{C}_{fb,n}^{noisy} = \frac{1}{n} \max_{Q(x^n||z^{n-1})} \max_{s_0} I(X^n \rightarrow Y^n||Z^{n-1}|s_0).$$

and

$$\bar{C}_{fb}^{noisy} \triangleq \lim_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy},$$

a limit that will be shown to exist in Theorem 43. In this section, we wish to show the following theorem.

**Theorem 39** *The capacity of an FSC with noisy feedback, if it exists, is upper bounded by*

$$C_{fb}^{noisy} \leq \bar{C}_{fb}^{noisy} \leq \left[ \bar{C}_{fb,n}^{noisy} + \frac{\log |\mathcal{S}|}{n} \right].$$

for all  $n > 0$ .

Before moving to the proof of this theorem, we present some necessary lemmas as follows. The omitted proofs of the following lemmas are given in Appendix.

#### 4.5.1 Necessary Technical Lemmas

**Lemma 40** *Let  $X^n, Y^n, Z^{n-1}$  be arbitrary random vector and  $S$  be a random variable taking values in an alphabet of size  $|\mathcal{S}|$ . Then*

$$|I(X^n \rightarrow Y^n|Z^{n-1}) - I(X^n \rightarrow Y^n||Z^{n-1}|S)| \leq H(S) \leq \log |\mathcal{S}|.$$



**Lemma 41** *For an FSC with noisy feedback as shown in Fig. 4.2, we have*

$$p(y_N|x^N, y^{N-1}, z^{N-1}, s^n) = p(y_N|x_{n+1}^N, y_{n+1}^{N-1}, z_n^{N-1}, s_n)$$

for all  $n \leq N - 1$ .

We next present the necessary sub-additive lemma below, the proof of which can be found in Appendix 4A Gallager (1968).

**Lemma 42** (*Sub-additive Sequence*) *Let  $a_N$ ,  $N = 1, 2, \dots, \infty$  be a bounded sequence of numbers. Assume that, for all  $1 \leq n < N$ ,*

$$Na_N \leq na_n + (N - n)a_{N-n}$$

then

$$\lim_{N \rightarrow \infty} a_N = \inf_N a_N.$$

#### 4.5.2 Proof of Theorem 39

First of all, we show a theorem as follows.

**Theorem 43** *For a noisy feedback FSC with  $|\mathcal{S}|$  states,*

$$\lim_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy} = \inf_n \left[ \bar{C}_{fb,n}^{noisy} + \frac{\log |\mathcal{S}|}{n} \right]$$

**Proof.** Let  $Q(x^N||z^{N-1})$  and  $s_0$  be the input distribution and the initial state that achieves  $\bar{C}_{fb,N}^{noisy}$ . All the distributions used in the following lines are determined by  $Q(x^N||z^{N-1})$  and

the channels:

$$\begin{aligned}
N\bar{C}_{fb,N}^{noisy} &= I(X^n \rightarrow Y^n || Z^{n-1} | s_0) \\
&= \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}, Z^{i-1}, s_0) + \sum_{i=n+1}^N I(X^i; Y_i | Y^{i-1}, Z^{i-1}, s_0) \\
&\leq n\bar{C}_{fb,n}^{noisy} + \sum_{i=n+1}^N I(X^i; Y_i | Y^{i-1}, Z^{i-1}, s_0) \\
&\stackrel{(a)}{\leq} n\bar{C}_{fb,n}^{noisy} + \sum_{i=n+1}^N I(X^i; Y_i | Y^{i-1}, Z^{i-1}, S_n, s_0) + \log |\mathcal{S}| \\
&= n\bar{C}_{fb,n}^{noisy} + \sum_{i=n+1}^N H(Y_i | Y^{i-1}, Z^{i-1}, S_n, s_0) - H(Y_i | X^i, Y^{i-1}, Z^{i-1}, S_n, s_0) + \log |\mathcal{S}| \\
&\stackrel{(b)}{\leq} n\bar{C}_{fb,n}^{noisy} + \sum_{i=n+1}^N H(Y_i | Y_{n+1}^{i-1}, Z_{n+1}^{i-1}, S_n) - H(Y_i | X_{n+1}^i, Y_{n+1}^{i-1}, Z_{n+1}^{i-1}, S_n) + \log |\mathcal{S}| \\
&= n\bar{C}_{fb,n}^{noisy} + \sum_{i=n+1}^N I(X_{n+1}^i; Y_i | Y_{n+1}^{i-1}, Z_{n+1}^{i-1}, S_n) + \log |\mathcal{S}| \\
&= n\bar{C}_{fb,n}^{noisy} + I(X_{n+1}^N \rightarrow Y_{n+1}^N || Z_{n+1}^{N-1} | S_n) + \log |\mathcal{S}| \\
&\leq n\bar{C}_{fb,n}^{noisy} + \max_{s_n} I(X_{n+1}^N \rightarrow Y_{n+1}^N || Z_{n+1}^{N-1} | s_n) + \log |\mathcal{S}| \\
&\stackrel{(c)}{\leq} n\bar{C}_{fb,n}^{noisy} + l\bar{C}_{fb,l}^{noisy} + \log |\mathcal{S}|
\end{aligned}$$

where line (a) is due to Lemma 40 where the first element in the sequence  $X^N$  is  $X^n$ . Inequality (b) follows from the fact that conditioning reduces entropy and the use of Lemma 41. Rearrange the last inequality, we have

$$N \left[ \bar{C}_{fb,N}^{noisy} + \frac{\log |\mathcal{S}|}{N} \right] \leq n \left[ \bar{C}_{fb,n}^{noisy} + \frac{\log |\mathcal{S}|}{n} \right] + l \left[ \bar{C}_{fb,l}^{noisy} + \frac{\log |\mathcal{S}|}{l} \right].$$

According to the sub-additive lemma 42, and since

$$\lim_{N \rightarrow \infty} \left[ \bar{C}_{fb,N}^{noisy} + \frac{\log |\mathcal{S}|}{N} \right] = \lim_{N \rightarrow \infty} \bar{C}_{fb,N}^{noisy},$$

the proof is complete. ■

Now, we are ready to show the proof of the main result. For reader's convenience, we recall the main result below and then present the proof.

**Theorem 44** *The capacity of an FSC with noisy feedback, if exists, is upper bounded by*

$$C_{fb}^{noisy} \leq \bar{C}_{fb}^{noisy} \leq \left[ \bar{C}_{fb,n}^{noisy} + \frac{\log |\mathcal{S}|}{n} \right].$$

**Proof.** Consider a code  $(n, 2^{nR})$  with average error probability  $P_e^{(n)}$ , we have

$$\begin{aligned}
NR &= H(M) \\
&= I(M; Y^N) + H(M|Y^N) \\
&\stackrel{(a)}{\leq} I(M; Y^N) + 1 + NP_e^{(N)}R \\
&\stackrel{(b)}{\leq} I(X^n \rightarrow Y^n || Z^{n-1}) + 1 + NP_e^{(N)}R \\
&\stackrel{(c)}{\leq} I(X^n \rightarrow Y^n || Z^{n-1}|S_0) + \log |\mathcal{S}| + 1 + NP_e^{(N)}R \\
&= \sum_{s_0 \in \mathcal{S}} p(s_0) I(X^n \rightarrow Y^n || Z^{n-1}|s_0) + \log |\mathcal{S}| + 1 + NP_e^{(N)}R \\
&\leq \max_{s_0} I(X^n \rightarrow Y^n || Z^{n-1}|s_0) + \log |\mathcal{S}| + 1 + NP_e^{(N)}R
\end{aligned}$$

where line (a) follows from Fano's inequality. Line (b) follows from Lemma 35 by replacing  $F^n$  by  $M$ . Line (c) is due to Lemma 40. By dividing both sides of the inequality by  $N$  and then take the limit, we have

$$\begin{aligned}
R &\leq \lim_{N \rightarrow \infty} \frac{1}{N} \max_{s_0} I(X^n \rightarrow Y^n || Z^{n-1}|s_0) + \log |\mathcal{S}| + 1 + NP_e^{(N)}R \\
&= \bar{C}_{fb}^{noisy} \\
&\leq \left[ \bar{C}_{fb,n}^{noisy} + \frac{\log |\mathcal{S}|}{n} \right]
\end{aligned}$$

where the last inequality follows from Theorem 43. ■

## 4.6 Conclusions

In this chapter, we studied the capacity of finite-alphabet channels with noisy feedback. We characterized the capacity in terms of the residual directed information by invoking the code-function representation used in Tatikonda and Mitter (2009). Then we provided a pair of upper and lower bounds on the capacity, which are characterized in terms of the causal conditioning directed information. Finally, we investigated finite state channels with noisy feedback and provided a finite dimensional optimization upper bound on the capacity.

## CHAPTER 5. BOUNDS ON THE CAPACITY OF GAUSSIAN CHANNELS WITH NOISY FEEDBACK

In this chapter, we turn our attention to Gaussian channels with additive Gaussian noisy feedback.

### 5.1 Introduction

We consider a discrete-time Gaussian channel with additive Gaussian noise feedback as shown in Fig.5.1. The additive Gaussian channel is modeled as

$$Y_i = X_i + W_i \quad i = 1, 2, \dots$$

where the gaussian noise  $\{W_i\}_{i=1}^\infty$  satisfies  $W^n = \{W_1, W_2, \dots, W_n\} \sim \mathbb{N}_n(0, \mathbf{K}_{w,n})$  for all  $n \in \mathbf{Z}^+$ . Similarly, the additive Gaussian feedback is modeled as

$$Z_i = Y_i + V_i \quad i = 1, 2, \dots$$

where the gaussian noise  $\{V_i\}_{i=1}^\infty$  satisfies  $V^n = \{V_1, V_2, \dots, V_n\} \sim \mathbb{N}_n(0, \mathbf{K}_{v,n})$  for all  $n \in \mathbf{Z}^+$ . Noise  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$  are assumed to be independent and  $\mathbf{K}_{w,n}$ ,  $\mathbf{K}_{v,n}$  are assumed to be nonsingular. Notice that we do not assumed stationarity of  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$ . For a code of rate  $R_n$  and length  $n$ , we specify a  $(n, 2^{nR_n})$  channel code as follows.  $M$  is an uniformly distributed message index where  $M \in \{1, 2, 3, \dots, 2^{nR_n}\}$ . There exists an encoding process  $X_i(M, Z^{i-1})$  for  $i = 1, 2, \dots, n$  ( $X_1(M, Z^0) = X_1(M)$ ), with power constraint

$$\frac{1}{n} \sum_{i=1}^n \mathbf{E} X_i^2(M, Z^{i-1}) \leq P,$$

and a decoding function  $g: Y^n \rightarrow \{1, 2, \dots, 2^{nR_n}\}$  with an error probability satisfying

$$P_e^{(n)} = \frac{1}{2^{nR_n}} \sum_{M=1}^{2^{nR_n}} p(M \neq g(y^n) | M) \leq \epsilon_n$$

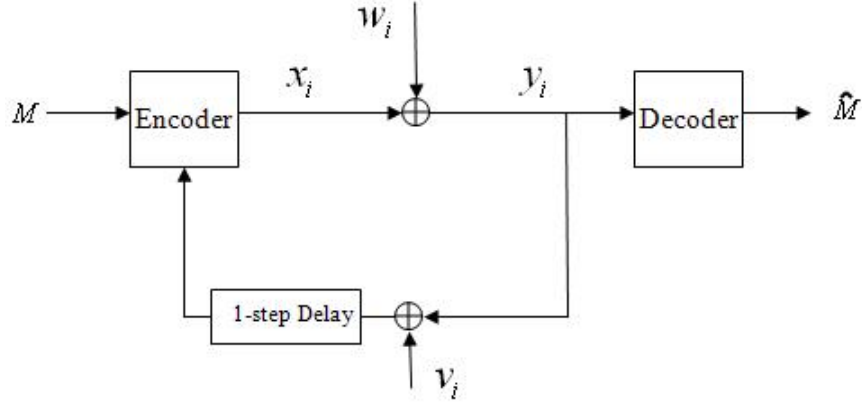


Figure 5.1 Gaussian channels with additive Gaussian noise feedback

where  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ . The objective of communication is to delivery  $M$  to the receiver at highest code rate with arbitrarily small error probability.

For time-varying Gaussian channels with noisy feedback, the standard notion of Shannon capacity may not exist. We thus define the  $n$ -block noisy feedback capacity  $C_{fb,n}^{noisy}$  by following Cover's definition (Theorem 1 in Cover and Pombra (1989)) on the  $n$ -block capacity of noiseless feedback Gaussian channels.

**Definition 45** (*N-block Noisy Feedback Capacity*)  $\{C_{fb,n}^{noisy}\}_{n=1}^{\infty}$  is an  $n$ -block noisy feedback capacity sequence if there exists a sequence of  $(n, 2^{n(C_{fb,n}^{noisy} - \epsilon)})$  codes with  $P_e^{(n)} \rightarrow 0$ , as  $n \rightarrow \infty$ , for  $\epsilon > 0$ ; conversely, for  $\epsilon > 0$ , any sequence of  $(n, 2^{n(C_{fb,n}^{noisy} + \epsilon)})$  codes has  $P_e^{(n)}$  bounded away from zero for all  $n$ .

Note that the above statements hold in the special cases of non-feedback and noiseless feedback with substitution of  $C_n$  (characterized in (5.1)) and  $C_{fb,n}$  (characterized in (5.3)) for  $C_{fb,n}^{noisy}$ , respectively. In sequel, “ $n$ -block capacity” is referred to “ $n$ -block noisy feedback capacity” for convenience, unless specified otherwise. We define the noisy feedback Shannon capacity

$$C_{fb}^{noisy} \triangleq \lim_{n \rightarrow \infty} C_{fb,n}^{noisy}$$

if the limit exists. Note that this definition of Shannon capacity is obviously the supremum of achievable rates and agrees with the conventional operational definition for the capacity of memoryless channels without feedback.

In this chapter, we wish to derive upper and lower bounds on the n-block capacity  $C_{fb,n}^{noisy}$  for arbitrary (stationary/nonstationary) Gaussian channels, and to find bounds on the Shannon capacity  $C_{fb}^{noisy}$  for stationary Gaussian channels.

In retrospect, additive Gaussian channels have been studied since the birth of “Information Theory” Shannon (1948). When there is no feedback (i.e.  $Z_i =$  for all  $i$ ), the channel input  $X_i$  is independent of the previous channel outputs. The n-block non-feedback capacity is characterized in Cover and Pombra (1989) as

$$C_n = \max_{tr(\mathbf{K}_{x,n}) \leq nP} \frac{1}{2n} \log \frac{\det(\mathbf{K}_{w,n} + \mathbf{K}_{x,n})}{\det \mathbf{K}_{w,n}} \quad (5.1)$$

where the maximum is taken over all positive semidefinite matrices  $\mathbf{K}_{x,n}$ . If we assume the stationarity on the process  $\{W_i\}_{i=1}^\infty$ , it is well-known that the nonfeedback (Shannon) capacity is characterized by water-filling on the noise power spectrum. Specifically,

$$C = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \frac{\max\{\mathbb{S}_w(e^{i\theta}), \lambda\}}{\mathbb{S}_w(e^{i\theta})} d\theta \quad (5.2)$$

where  $\mathbb{S}_w(e^{i\theta})$  is the power spectrum density of the stationary noise process  $\{W_i\}_{i=1}^\infty$ . The water level  $\lambda$  should satisfy

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \max\{0, \lambda - \mathbb{S}_w(e^{i\theta})\} d\theta = P.$$

Note that the initial idea of water-filling should be attributed to Shannon Shannon (1949).

When there is a noiseless feedback (i.e.  $Z_i = Y_{i-1}$  for all  $i$ ), the n-block feedback capacity is notably characterized in Cover and Pombra (1989) as

$$C_{fb,n} = \max_{\mathbf{B}_n, \mathbf{K}_{s,n}} \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n)\mathbf{K}_{w,n}(\mathbf{I}_n + \mathbf{B}_n)^T + \mathbf{K}_{s,n})}{\det \mathbf{K}_{w,n}} \quad (5.3)$$

where the maximum is taken over all positive semidefinite matrices  $\mathbf{K}_{s,n}$  and all strictly lower triangular matrices  $\mathbf{B}_n$  satisfying

$$\frac{1}{n} tr(\mathbf{K}_{s,n} + \mathbf{B}_n \mathbf{K}_{w,n} \mathbf{B}_n^T) \leq P. \quad (5.4)$$

Similar to the nonfeedback case, if we assume the stationarity on the process  $\{W_i\}_{i=1}^\infty$ , the noiseless feedback (Shannon) capacity is characterized in Kim (2010) as

$$C_{fb} = \sup_{\mathbb{S}_s, \mathbb{B}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \frac{\mathbb{S}_s(e^{i\theta}) + |1 + \mathbb{B}(e^{i\theta})|^2 \mathbb{S}_w(e^{i\theta})}{\mathbb{S}_w(e^{i\theta})} d\theta, \quad (5.5)$$

with power constraint

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbb{S}_s(e^{i\theta}) + |\mathbb{B}(e^{i\theta})|^2 \mathbb{S}_w(e^{i\theta}) d\theta \leq P. \quad (5.6)$$

Here  $\mathbb{B}(e^{i\theta})$  represents all possible strictly causal linear filters. However, when there is an additive Gaussian noise feedback as shown in Fig.5.1, the characterizations on the n-block capacity  $C_{fb,n}^{noisy}$  or the capacity  $C_{fb}^{noisy}$  under stationarity assumption on the Gaussian noise has not been developed yet, to the present author's knowledge.

So far, only few papers have addressed noisy Gaussian feedback problem or its variations. Chance and Love (2010) and Li and Elia (2011a) took Cover-Pombra scheme into account for noisy feedback Gaussian channels and derive the upper and lower bounds on its maximal achievable rate. Other works focus on the additive white Gaussian noise (AWGN) channel with AWGN feedback, whose capacity is known to be the nonfeedback capacity. For example, Kim et al. (2007) derived the upper and lower bounds on the reliability function and shows that, noise in the feedback link renders the noisy feedback communication fundamentally different from the noiseless feedback case. Wyner (1969), Martins and Weissman (2008) and Chance and Love (2011a) proposed specific coding/decoding schemes based on the notable Schalkwijk-Kailath Scheme (Schalkwijk and Kailath (1966)).

In this chapter, we derive a pair of n-block upper and lower bounds, denoted as  $\bar{C}_{fb,n}^{noisy}$  and  $\underline{C}_{fb,n}^{noisy}$ , on the n-block capacity  $C_{fb,n}^{noisy}$ . The main feature of these n-block bounds is that they can be characterized as convex optimization problems and obtained efficiently by using standard optimization tools (e.g. semi-definite programming). Next, we consider stationary Gaussian channels with stationary Gaussian noise feedback, for which the Shannon capacity  $C_{fb}^{noisy}$  may exist, and if  $C_{fb}^{noisy}$  exists, we have

$$\liminf_{n \rightarrow \infty} C_{fb,n}^{noisy} \leq C_{fb}^{noisy} \leq \limsup_{n \rightarrow \infty} C_{fb,n}^{noisy}. \quad (5.7)$$

Applying the proposed n-block upper and lower bounds, we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \underline{C}_{fb,n}^{noisy} &\leq \liminf_{n \rightarrow \infty} C_{fb,n}^{noisy} \leq C_{fb}^{noisy} \\ &\leq \limsup_{n \rightarrow \infty} C_{fb,n}^{noisy} \leq \limsup_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy}. \end{aligned} \quad (5.8)$$

Following Kim's approach on the stationary Gaussian channels with noiseless feedback Kim (2010), we show that the limits of the n-block upper and lower bounds exist and can be characterized as a power spectral optimization problem (i.e. a single infinite dimensional optimization problem). However, solving these power spectral optimization problems is notoriously difficult in general. Then we develop some results to evaluate/bound these characterizable limit values of  $\bar{C}_{fb,n}^{noisy}$  and  $\underline{C}_{fb,n}^{noisy}$ . Next, we use two approaches to obtain a lower bound on the  $\liminf_{n \rightarrow \infty} \underline{C}_{fb,n}^{noisy}$ , which clearly is a lower bound on Shannon capacity  $C_{fb}^{noisy}$ : 1) By using superadditive property of a sequence, we prove that for all  $n \geq 1$ ,

$$\underline{C}_{fb,n}^{noisy} \leq \lim_{n \rightarrow \infty} \underline{C}_{fb,n}^{noisy}, \quad (5.9)$$

2) We propose a control-oriented (LQG control) linear coding scheme for a new noiseless feedback Gaussian channel whose achievable rate is guaranteed to be a lower bound on  $C_{fb}^{noisy}$ .

To end this section, we present some preliminary results which will be used later in this chapter.

Firstly, the entropy-maximization lemma is presented as follows.

**Lemma 46** (*Cover and Thomas (2006)*) *Let the random vector  $X^n \in \mathbb{R}^n$  have zero mean and covariance  $\mathbf{K}_{x,n} = \mathbb{E}X^n X^{nT}$  (i.e.  $\mathbf{K}_{x,n}(i, j) = \mathbb{E}X_i X_j$ ,  $1 \leq i, j \leq n$ ). Then*

$$h(X^n) \leq \frac{1}{2} \log(2\pi e)^n \det \mathbf{K}_{x,n}$$

*with equality if and only if  $X^n \sim \mathbb{N}_n(0, \mathbf{K}_{x,n})$ .*

We next recall the Schur complement. We refer the interested readers to Appendix A in Boyd and Vandenberghe (2004) for a comprehensive introduction on the Schur complement decomposition.

**Definition 47** (*Schur Complement*) *Consider an  $n \times n$  symmetric matrix  $X$  partitioned as*

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}.$$



If  $\det A \neq 0$ , the matrix

$$S = C - B^T A^{-1} B$$

is called the Schur complement of  $A$  in  $X$ .

We present some properties of the Schur complement as follows.

1.  $\det X = \det A \cdot \det S$ .
2.  $X \succ 0$  if and only if  $A \succ 0$  and  $S \succ 0$ .
3. If  $A \succ 0$ , then  $X \succeq 0$  if and only if  $S \succeq 0$ .

## 5.2 Information Flow Revisited

With the ultimate long-term objective to find the capacity of the noisy feedback channels, it is of importance to understand how extraneous information (e.g. message, additive noise) is delivered to the decoder through the forward channel and how they interfere with each other. In this section, we revisit the information flow decomposition equality, which unveils the information flow pattern in additive Gaussian channels with additive Gaussian noise feedback. As a natural application of this decomposition, we next derive an upper bound on  $I(M; Y^n)$ , the superior limit of which is an upper bound on the noisy feedback capacity according to Fano's inequality. Now, we recall a result, Corollary 22, in Chapter 3.

**Corollary 48** *For additive Gaussian channels with additive Gaussian noise feedback, as shown in Fig.5.1,*

$$I(X^n \rightarrow Y^n) = I(M; Y^n) + I(V^{n-1}; Y^n) + I(M; V^{n-1} | Y^n).$$

With this corollary in hand, we now derive an upper bound on the n-block capacity. Consider a channel code with rate  $R_n$  and length  $n$ , we begin with Fano's inequality on the entropy  $H(M | Y^n)$ , specifically,

$$\begin{aligned} nR_n &= H(M) \\ &= H(M | Y^n) + I(M; Y^n) \\ &\leq I(M, Y^n) + n\epsilon_n \end{aligned}$$

where  $\epsilon_n \rightarrow 0$  if  $Pe^{(n)} \rightarrow 0$ . As  $R_n$  refers to any achievable rate, this inequality implies an upper bound,  $\frac{1}{n} \max I(M; Y^n)$ , on the n-block capacity  $C_{fb,n}^{noisy}$ . However, how to characterize  $I(M; Y^n)$  in a noisy feedback setting is still unknown. Now, based on Corollary 22, we have

$$\begin{aligned} I(M; Y^n) &= I(X^n \rightarrow Y^n) - I(V^{n-1}; Y^n) - I(M; V^{n-1} | Y^n) \\ &\leq I(X^n \rightarrow Y^n) - I(V^{n-1}; Y^n), \end{aligned}$$

which indicates that  $I(X^n \rightarrow Y^n) - I(V^{n-1}; Y^n)$  is an upper bound on  $I(M; Y^n)$ . The next theorem shows that this upper bound can be represented in single term,  $I(X^n \rightarrow Y^n | V^{n-1})$ , instead of the subtraction of two information quantities. Clearly,  $\frac{1}{n} \max I(X^n \rightarrow Y^n | V^{n-1})$  is an upper bound on the n-block capacity.

**Theorem 49** *For additive Gaussian channels with additive Gaussian noise feedback, as shown in Fig.5.1,*

$$I(M; Y^n) = I(X^n \rightarrow Y^n | V^{n-1}) - I(M; V^{n-1} | Y^n)$$

Furthermore, we have

$$I(X^n \rightarrow Y^n | V^{n-1}) = h(Y^n | V^{n-1}) - h(W^n).$$

**Remark 50** *Two main advantages of using  $I(X^n \rightarrow Y^n | V^{n-1})$  as an upper bound on  $I(M; Y^n)$  are discussed below,*

1. *The value of  $I(X^n \rightarrow Y^n | V^n)$  only depends on channel input distributions  $\{f(x_i | x^{i-1}, y^{i-1} + v^{i-1})\}_{i=1}^\infty$  instead of particular codewords, which might significantly reduce the computation complexity.*
2. *Based on the equality*

$$I(X^n \rightarrow Y^n | V^{n-1}) = I(X^n \rightarrow Y^n) - I(V^{n-1}; Y^n),$$

*we clearly have*

$$I(M; Y^n) \leq I(X^n \rightarrow Y^n | V^{n-1}) \leq I(X^n \rightarrow Y^n).$$

This indicates that  $I(X^n \rightarrow Y^n | V^{n-1})$  is a better upper bound than  $I(X^n \rightarrow Y^n)$ , which is widely used in characterizing the capacity of noiseless feedback channels. Note that in noiseless feedback case,  $I(M; Y^n) = I(X^n \rightarrow Y^n)$ .

To prove this theorem, we need a lemma as follows.

**Lemma 51** For Gaussian channels with Gaussian noise feedback as shown in Fig. 5.1, we have

1.  $h(Y^n | V^{n-1}) = h(Y^n | V^{n-1})$
2.  $I(X^n \rightarrow Y^n | V^{n-1}) = I(X^n \rightarrow Y^n | V^{n-1})$

**Proof.** See Appendix. ■

Now, we show the proof of Theorem 49.

**Proof.** First of all, we have

$$\begin{aligned}
 & I(X^n \rightarrow Y^n | V^{n-1}) \\
 &= \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}, V^{i-1}) \\
 &= \sum_{i=1}^n h(Y_i | Y^{i-1}, V^{i-1}) - h(Y_i | X^{i-1}, Y^{i-1}, V^{i-1}) \\
 &= h(Y^n | V^{n-1}) - \sum_{i=1}^n h(Y_i | X^{i-1}, Y^{i-1}, V^{i-1}) \\
 &= h(Y^n | V^{n-1}) - \sum_{i=1}^n h(Y_i | X^{i-1}, Y^{i-1}, V^{i-1})
 \end{aligned}$$

where the last line uses Lemma 51 (1). Next,

$$\begin{aligned}
 & I(X^n \rightarrow Y^n | V^{n-1}) \\
 &= \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}, V^{n-1}) \\
 &= \sum_{i=1}^n h(Y_i | Y^{i-1}, V^{n-1}) - h(Y_i | X^{i-1}, Y^{i-1}, V^{n-1}) \\
 &= h(Y^n | V^{n-1}) - \sum_{i=1}^n h(Y_i | X^{i-1}, Y^{i-1}, V^{n-1})
 \end{aligned}$$

Based on Lemma 51(2), it is straightforward to have

$$\sum_{i=1}^n h(Y_i|X^{i-1}, Y^{i-1}, V^{i-1}) = \sum_{i=1}^n h(Y_i|X^{i-1}, Y^{i-1}, V^{n-1}).$$

Now, we are ready to show

$$I(X^n \rightarrow Y^n) - I(V^{n-1}; Y^n) = I(X^n \rightarrow Y^n|V^{n-1}).$$

That is,

$$\begin{aligned} & I(X^n \rightarrow Y^n) - I(V^{n-1}; Y^n) \\ &= \sum_{i=1}^n I(X^i; Y_i|Y^{i-1}) - I(V^{n-1}; Y^n) \\ &= \sum_{i=1}^n h(Y_i|Y^{i-1}) - h(Y_i|X^i, Y^{i-1}) - I(V^{n-1}; Y^n) \\ &= \sum_{i=1}^n h(Y_i|Y^{i-1}) - h(Y_i|X^i, Y^{i-1}, V^{i-1}) - I(V^{n-1}; Y^n) \\ &\stackrel{(a)}{=} h(Y^n) - \sum_{i=1}^n h(Y_i|X^i, Y^{i-1}, V^{n-1}) - h(Y^n) + h(Y^n|V^{n-1}) \\ &= \sum_{i=1}^n h(Y_i|Y^{i-1}, V^{n-1}) - \sum_{i=1}^n h(Y_i|X^i, Y^{i-1}, V^{n-1}) \\ &= \sum_{i=1}^n I(X^i, Y_i|Y^{i-1}, V^{n-1}) \\ &= I(X^n \rightarrow Y^n|V^{n-1}) \end{aligned}$$

where line (a) follows from equation (5.2). Next, based on the line before line (a), we have

$$\begin{aligned} & I(X^n \rightarrow Y^n|V^{n-1}) \\ &= \sum_{i=1}^n h(Y_i|Y^{i-1}) - h(Y_i|X^i, Y^{i-1}, V^{i-1}) - I(V^{n-1}; Y^n) \\ &= \sum_{i=1}^n h(Y_i|Y^{i-1}, V^{n-1}) - h(Y_i|X^i, Y^i, V^{i-1}) \\ &\stackrel{(b)}{=} h(Y^n|V^{n-1}) - h(X_i + W_i|W^{i-1}, X^i, Y^{i-1}, V^{i-1}) \\ &\stackrel{(c)}{=} h(Y^n|V^{n-1}) - h(W_i|W^{i-1}) \\ &= h(Y^n|V^{n-1}) - h(W^n) \end{aligned}$$

where line (b) follows from  $Y_i = X_i + W_i$ , and line (c) follows from the Markov chain  $W_i - W^{i-1} - (X^i, Y^{i-1}, V^{i-1})$  which holds due to the causality of these random variables. ■

### 5.3 Upper Bound Characterization

In light of Theorem 49, in this section, we characterize an upper bound  $\bar{C}_{fb,n}^{noisy}$  on the  $n$ -block capacity  $C_{fb,n}^{noisy}$  for arbitrary Gaussian channels with noisy feedback. It is shown that  $\bar{C}_{fb,n}^{noisy}$  can be obtained by solving a convex optimization problem. Next, we consider stationary Gaussian channels for which the Shannon capacity may exist. We characterize  $\limsup_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy}$  as an infinite dimensional optimization problem, providing an upper bound on the Shannon capacity, if it exists. We then provide some preliminary results which are helpful to numerically evaluate  $\limsup_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy}$ .

#### 5.3.1 Upper Bound On The N-block Capacity

In this subsection, we wish to show the following Theorem.

**Theorem 52** *Consider an additive Gaussian noise channel with additive Gaussian noise feedback as shown in Fig.5.1. An upper bound  $\bar{C}_{fb,n}^{noisy}$  on the  $n$ -block capacity (Definition 45) can be obtained by solving*

$$\begin{aligned} & \underset{\mathbf{B}_n, \mathbf{K}_{s,n}}{\text{maximize}} \quad \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n)\mathbf{K}_{w,n}(\mathbf{I}_n + \mathbf{B}_n)^T + \mathbf{K}_{s,n})}{\det \mathbf{K}_{w,n}} \\ & \text{subject to} \quad \text{tr}(\mathbf{K}_{s,n} + \mathbf{B}_n(\mathbf{K}_{v,n} + \mathbf{K}_{w,n})\mathbf{B}_n^T) \leq nP \\ & \quad \mathbf{K}_{s,n} \geq 0, \quad \mathbf{B}_n \text{ is strictly lower triangular.} \end{aligned} \tag{5.10}$$

**Remark 53** *We see that 5.10 is a generalization of the well know formula for noiseless feedback. Compared with noiseless feedback formula (5.3), we find that the feedback noise covariance  $\mathbf{K}_{v,n}$  only appears in the power constraint. When  $\mathbf{K}_{v,n}$  is small (in the positive semi-definite cone), the  $n$ -block capacity intuitively converges to the  $n$ -block noiseless feedback capacity. As it is shown in (5.10), this proposed  $n$ -block upper bound also converges to the  $n$ -block noiseless feedback capacity and, therefore, the upper bound should be tight. When  $\mathbf{K}_{v,n}$  is large (in the positive semi-definite cone), the matrix  $\mathbf{B}_n$  would be close to a zero matrix. This implies the feedback is almost “shut-off” (to be seen from the simulation results) and this upper bound characterization converges to the  $n$ -block non-feedback capacity characterization (formula (5.1)). Thus, this bound is tight in the regime of both small and large feedback noise  $V$ .*

The next result shows that the above optimization problem can be transformed into a convex form which can be efficiently solved by the standard semidefinite programming.

**Corollary 54** *The upper bound  $\bar{C}_{fb,n}^{noisy}$  on the  $n$ -block noisy feedback capacity can be obtained by solving the following convex optimization problem,*

$$\begin{aligned} & \underset{\mathbf{H}_n, \mathbf{B}_n}{\text{maximize}} && \frac{1}{2n} \log \det \begin{bmatrix} \mathbf{K}_{v,n}^{-1} & \mathbf{B}_n^T \\ \mathbf{B}_n & \mathbf{H}_n \end{bmatrix} - \frac{1}{2n} \log \det(\mathbf{K}_{v,n}^{-1} \mathbf{K}_{w,n}) \\ & \text{subject to} && \text{tr}(\mathbf{H}_n - \mathbf{K}_{w,n} \mathbf{B}_n^T - \mathbf{B}_n \mathbf{K}_{w,n} - \mathbf{K}_{w,n}) \leq nP \\ & && \begin{bmatrix} \mathbf{H}_n & \mathbf{I}_n + \mathbf{B}_n^T & \mathbf{B}_n^T \\ \mathbf{I}_n + \mathbf{B}_n & \mathbf{K}_{w,n}^{-1} & \mathbf{0}_n \\ \mathbf{B}_n & \mathbf{0}_n & \mathbf{K}_{v,n}^{-1} \end{bmatrix} \geq 0 \\ & && \mathbf{B}_n \text{ is strictly lower triangular.} \end{aligned}$$

The complexity of computing the above convex optimization problem is evaluated in the following proposition.

**Proposition 55** *The complexity of solving the linear matrix inequality(LMI) optimization problem in Corollary 54 is upper bounded by*

$$\mathcal{O}\left(\frac{81}{8}n^7 - \frac{27}{4}n^6 + \frac{3}{4}n^4 - \frac{1}{8}n^3\right).$$

**Proof.** See Appendix. ■

In the rest of this subsection, we show the proof of Theorem 52. The basic idea of the proof is as follows. Based on Theorem 49 and Fano's inequality, it is known that  $\frac{1}{n} \max I(X^n \rightarrow Y^n | V^{n-1})$  is an upper bound on the  $n$ -block noisy feedback capacity  $C_{fb,n}^{noisy}$  where the max is taken over all admissible coding schemes<sup>1</sup>. We propose a Cover-Pombra(CP)-like coding scheme and show that the maximization over the CP-like scheme does not lose optimality. Then characterize  $\frac{1}{n} \max I(X^n \rightarrow Y^n | V^{n-1})$  under CP-like scheme results in formula (5.10).

---

<sup>1</sup>“admissible” means that the power constraint is satisfied.

### 5.3.1.1 Cover-Pombra(CP)-like Coding Scheme

The CP-like scheme consists of linear encoding of the feedback signal and Gaussian signaling of the message, as shown in a vector form in Fig.5.2. Specifically,

$$\text{The channel input signal: } X^n = S^n + \mathbf{B}_n(W^n + V^n),$$

$$\text{The channel output signal: } Y^n = S^n + \mathbf{B}_n(W^n + V^n) + W^n,$$

$$\text{The power constraint: } \text{tr}(\mathbf{K}_{s,n} + \mathbf{B}_n(\mathbf{K}_{w,n} + \mathbf{K}_{v,n})\mathbf{B}_n^T) \leq nP,$$

where  $S^n = \{S_1, S_2, \dots, S_n\} \sim \mathbb{N}_n(0, \mathbf{K}_{s,n})$  is the message information vector and independent from  $V^n$  and  $W^n$ .  $\mathbf{B}_n$  is an  $n \times n$  strictly lower triangular linear encoding matrix. Note that the one-step delay in the feedback link is captured by the particular structure of matrix  $\mathbf{B}_n$ .

**Remark 56** *This CP-like coding scheme can be specifically expressed as a concatenated coding scheme as shown in Fig. 5.3, which can be verified by checking the channel inputs and outputs. The outer encoder  $\mathbf{E}_1$  maps each message index to a vector  $s^n$  which is drawn from the distribution  $\mathbb{N}_n(0, \mathbf{K}_{s,n})$ . The inner encoder linearly takes the message information vector and the feedback information to produce channel inputs.*

### 5.3.1.2 Proof of Theorem 52

Now we show the proof of Theorem 52, that is, any  $(n, 2^{n(\bar{C}_{fb,n}^{noisy} + \epsilon)})$  noisy feedback channel codes have probability of error  $P_e^{(n)}$  bounded away from zero.

**Proof.** According to Definition 45, we wish to show that a sequence of  $(n, 2^{nR_n})$  channel

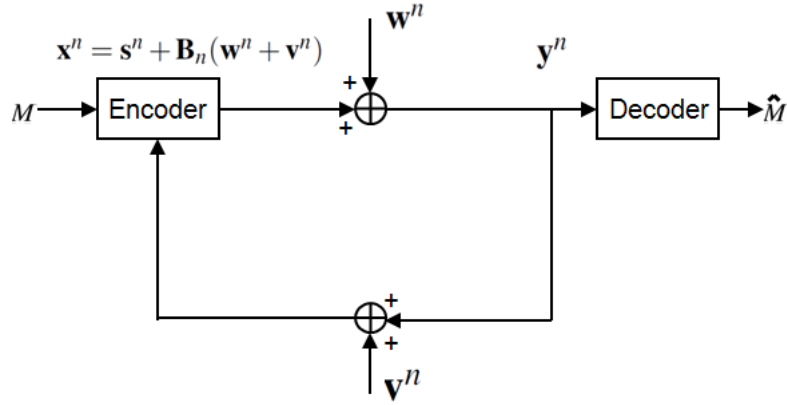


Figure 5.2 Gaussian channels with additive Gaussian noise feedback(Gaussian signalling and linear feedback)

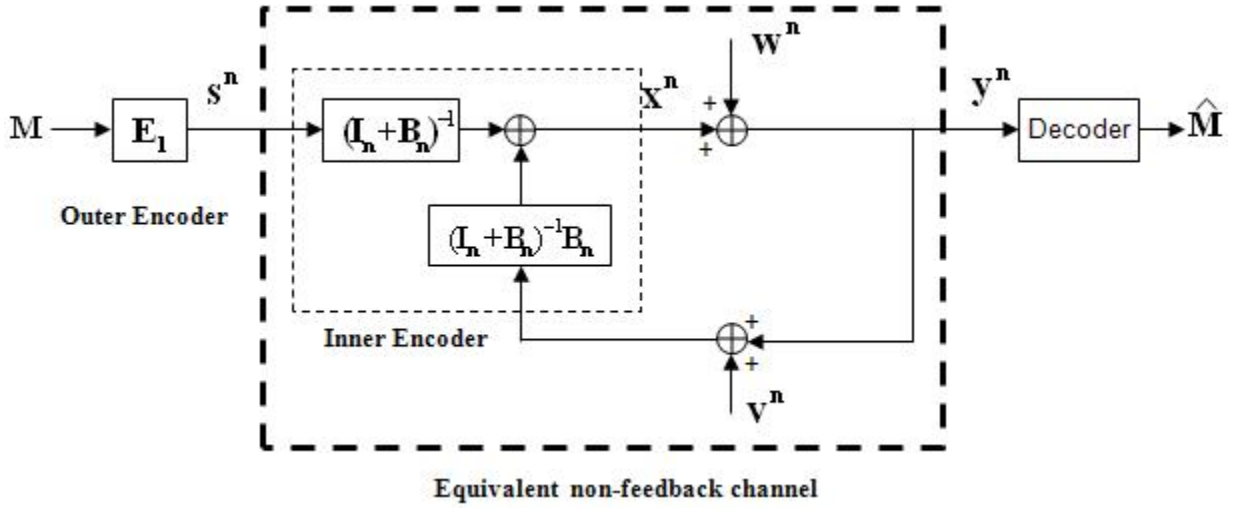


Figure 5.3 A concatenated coding representation of CP-like Scheme. The inner linear encoder can be also interpreted as a portion of the equivalent non-feedback channel.



codes with  $Pe^{(n)} \rightarrow 0$  must have  $R_n \leq \bar{C}_{fb,n}^{noisy} + \delta_n$  where  $\delta_n \rightarrow 0$ . By Fano's inequality,

$$\begin{aligned} nR_n &= H(M) \\ &= H(M|Y^n) + I(M; Y^n) \\ &= I(M, Y^n) + n\delta_n \end{aligned}$$

where  $\delta_n \rightarrow 0$  if  $Pe^{(n)} \rightarrow 0$ . Then, by Theorem 49,

$$\begin{aligned} nR_n &= I(X^n \rightarrow Y^n | V^{n-1}) - I(M; V^n | Y^n) + n\delta_n \\ &\leq I(X^n \rightarrow Y^n | V^{n-1}) + n\delta_n \\ &= h(Y^n | V^{n-1}) - h(W^n) + n\delta_n \\ &\leq \max_{\text{all admissible coding schemes}} h(Y^n | V^{n-1}) - h(W^n) + \delta_n \end{aligned}$$

Denote

$$\bar{C}_{fb,n}^{noisy} = \max_{\text{all admissible coding schemes}} h(Y^n | V^{n-1}) - h(W^n). \quad (5.11)$$

We next show that maximizing  $h(Y^n | V^{n-1}) - h(W^n)$  over CP-like coding schemes (Fig. 5.2) does not lose the optimality. That is,

$$\begin{aligned} &\max_{\text{all admissible coding schemes}} \frac{1}{n} (h(Y^n | V^{n-1}) - h(W^n)) \\ &= \max_{\text{CP-like scheme}} \frac{1}{n} (h(Y^n | V^{n-1}) - h(W^n)) \end{aligned}$$

Since we can not affect the noise entropy (i.e.  $h(W^n)$ ), we need to maximize  $h(Y^n | V^{n-1})$  over all admissible channel inputs  $\{X_i\}_{i=1}^n$ . First of all, we have

$$h(Y^n | V^{n-1}) = h(Y^n, V^{n-1}) - h(V^{n-1}).$$

Based on Lemma 46, the random variables  $(Y^n, V^{n-1})$  should be jointly Gaussian in order to maximize  $h(Y^n, V^{n-1})$ . As  $V^{n-1}$  has Gaussian distribution,  $Y^n$  must be Gaussian. Furthermore, since  $W^n$  is Gaussian and  $Y^n = X^n + W^n$ , it suffices to construct  $X^n$  satisfying Gaussian distribution. Another fact is that  $X^n$  should depend on  $W^{n-1} + V^{n-1}$  instead of  $W^{n-1}$  and  $V^{n-1}$  individually since the channel outputs are fed back to the encoder without any encoding.

Therefore, the most general normal causal dependence of  $X^n$  on the previous feedback information  $Y^{n-1} + V^{n-1}$  satisfying the above arguments is in the form of  $X^n = S^n + \mathbf{B}_n(W^n + V^n)$

where  $\mathbf{B}_n$  is strictly lower-triangular matrix. Then we have

$$\begin{aligned}
& h(Y^n|V^{n-1}) - h(W^n) \\
&= h(S^n + \mathbf{B}_n(W^n + V^n) + W^n|V^{n-1}) - h(W^n) \\
&\stackrel{(a)}{=} h(S^n + (\mathbf{B}_n + \mathbf{I}_n)W^n|V^{n-1}) - h(W^n) \\
&\stackrel{(b)}{=} h(S^n + (\mathbf{B}_n + \mathbf{I}_n)W^n) - h(W^n)
\end{aligned}$$

where line (a) uses the fact, as  $\mathbf{B}_n$  is a strictly lower-triangular matrix,  $\mathbf{B}_n V^n$  only depends on  $V^{n-1}$ . Line (b) follows from the independence of  $S^n$  and  $W^n$  on  $V^{n-1}$ . By Lemma 46, it is straightforward to obtain

$$R_n \leq \bar{C}_{fb,n}^{noisy} + \delta_n,$$

which implies that

$$C_{fb,n}^{noisy} \leq \bar{C}_{fb,n}^{noisy} + \delta_n.$$

■

**Remark 57** *The CP-like coding scheme may not be the optimal (capacity-achieving) coding scheme for noisy feedback Gaussian channels. We herein adopt this coding scheme only because it can nicely characterize the proposed  $n$ -block upper bound. Note that the CP-like scheme may not apply if we look at a different upper bound.*

### 5.3.2 Characterization Under Stationary Gaussian Noises

In this subsection, we add stationarity assumption to additive Gaussian noises. Then the Shannon capacity  $C_{fb}^{noisy}$ , if it exists, is upper bounded as

$$C_{fb}^{noisy} \leq \limsup_{n \rightarrow \infty} C_{fb,n}^{noisy} \leq \limsup_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy}. \quad (5.12)$$

We show that  $\lim_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy}$  exists and can be characterized as a single infinite dimensional optimization problem.

**Theorem 58** *Assume that  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$  are stationary processes. Then the limit of the proposed  $n$ -block upper bound (5.10) exists and can be characterized as*

$$\bar{C}_{fb}^{noisy} = \sup_{\mathbb{S}_s, \mathbb{B}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \frac{\mathbb{S}_s(e^{i\theta}) + |1 + \mathbb{B}(e^{i\theta})|^2 \mathbb{S}_w(e^{i\theta})}{\mathbb{S}_w(e^{i\theta})} d\theta \quad (5.13)$$

with power constraint

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbb{S}_s(e^{i\theta}) + |\mathbb{B}(e^{i\theta})|^2 (\mathbb{S}_w(e^{i\theta}) + \mathbb{S}_v(e^{i\theta})) d\theta \leq P. \quad (5.14)$$

Here,  $\mathbb{S}_w(e^{i\theta})$  and  $\mathbb{S}_v(e^{i\theta})$  are the power spectral density of  $\{W_i\}_{i=1}^{\infty}$  and  $\{V_i\}_{i=1}^{\infty}$ , respectively, and the maximization is taken over all power spectral density  $\mathbb{S}_s(e^{i\theta}) \geq 0$  and strictly causal linear filter  $\mathbb{B}(e^{i\theta}) = \sum_{k=1}^{\infty} b_k e^{ik\theta}$ .

The proof invokes and extends Kim's approach in Kim (2010) to noisy feedback settings by replacing directed information by conditional directed information. We present the details of the proof in Appendix.

As it is shown, the limit value  $\bar{C}_{fb}^{noisy}$  is characterized by a single infinite dimensional optimization problem that is in general difficult to solve. This is not unexpected, as the characterization for noiseless feedback has similar difficulties and has been computed so far only for the 1st-order autoregressive moving average (ARMA(1)) stationary Gaussian channel model. In the noisy feedback setting, even this case is not easily computable. Next, we use Riemann approximation approach to evaluate  $\bar{C}_{fb}^{noisy}$ . Specifically, the region  $[-\pi, \pi]$  is divided into (sufficiently large)  $n$  equal partitions. For the  $j$ -th partition ( $j = 0, 1, 2, \dots, n$ ), we define

$$u_j = \log \frac{s_j + |1 + a_j + ib_j|^2 s_{w,j}}{s_{w,j}}$$

where  $s_j$ ,  $s_{w,j}$  and  $a_j + ib_j$  are the value of  $\mathbb{S}_s(e^{i\theta})$ ,  $\mathbb{S}_w(e^{i\theta})$  and  $\mathbb{B}(e^{i\theta})$  evaluated at  $\theta = \frac{2\pi}{n}j - \pi$ , respectively. Note that evaluate  $\mathbb{B}(e^{i\theta})$  by  $a_j + ib_j$  does not capture the causality of the filter  $\mathbb{B}(e^{i\theta})$ . Therefore, in order to have an accurate approximation we need to add causality constraints in terms of  $a_j$  and  $b_j$  on the filter. The initial attempt goes to the properness of the open-system and the induced property on the sensitivity function. Specifically, for “strictly proper” systems with feedback, we have the following fact,

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} |\mathcal{S}(e^{2\pi\theta})|^2 d\theta = 1 + \int_{-\frac{1}{2}}^{\frac{1}{2}} |\mathcal{T}(e^{2\pi\theta})|^2 d\theta \quad (5.15)$$

where  $\mathcal{S}$  and  $\mathcal{T}$  represent the sensitivity function and complementary sensitivity function, respectively. We refer the interested readers to the vast literature on the proof of this equality. By applying this equality to the CP-like feedback system described in 5.3.1.1 (See Fig.5.3), the

sensitivity function of the feedback CP-like system ( $n \rightarrow \infty$ ) is  $1 + \mathbb{B}$  and the complementary function is  $\mathbb{B}$ . According to the above equality and the notation  $\mathbb{B}(e^{i\theta}) = a(\theta) + ib(\theta)$ , we have

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} |1 + a(\theta) + ib(\theta)|^2 d\theta = 1 + \int_{-\frac{1}{2}}^{\frac{1}{2}} |a(\theta) + ib(\theta)|^2 d\theta. \quad (5.16)$$

After some algebra we have the following constraint

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} a(\theta) d\theta = 0. \quad (5.17)$$

By using Riemann approximation,

$$\sum_{j=1}^n a_j = 0. \quad (5.18)$$

Putting above together, the optimization problem (5.13) can be reformulated by using Riemann approximation as

$$U_{fb}^{noisy} = \sup_{s_j, a_j, b_j, j=1,2,\dots,n} \frac{1}{4\pi} \sum_{j=1}^n u_j \quad (5.19)$$

with power constraint

$$\begin{aligned} \sum_{j=1}^n a_j &= 0, \quad s_j > 0, \\ \frac{1}{2\pi} \sum_{j=1}^n s_j + |a_j + ib_j|^2 (s_{w,j} + s_{v,j}) &\leq P. \end{aligned} \quad (5.20)$$

where  $s_{v,j}$  is the value of  $\mathbb{S}_v(e^{i\theta})$  evaluated at  $\theta = \frac{2\pi}{n}j - \pi$ .

According to Wang and Elia (2011), the above optimization problem can be efficiently solved in a distributed fashion. Notice that the optimization problem 5.19 is not equivalent to the original problem (5.13) even with  $n \rightarrow \infty$ . This is because we do not fully capture the causality constraint on  $\mathbb{B}(e^{i\theta}) = \sum_{k=1}^{\infty} b_k e^{ik\theta}$  although we already consider the closed-loop constraint on causal feedback systems. Thus, the objective value  $U_{fb}^{noisy}$  with sufficient large  $n$  is an upper bound on  $\bar{C}_{fb}^{noisy}$ . A future research will focus on exploring the causality constraints (represented in the frequency domain) such that the approximation is as accurate as possible. A promising attempt will go to the Hilbert transform, which can be used to determine a system's causality from its frequency response. Whether this additional constraint will lead to a desired approximation on  $\bar{C}_{fb}^{noisy}$ , however, remains to be seen.

### 5.3.2.1 Properties of Sequence $\bar{C}_{n,fb}^{noisy}$

With the objective to obtain the exact value of  $\bar{C}_{fb}^{noisy}$ , in what follows, we provide some preliminary results on the properties of the sequence  $\bar{C}_{n,fb}^{noisy}$  ( $n = 1, 2, \dots$ ), which probably can be used to evaluate  $\bar{C}_{fb}^{noisy}$  in the future research.

**Lemma 59** *Assume the Gaussian noises  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$  are stationary. Then*

$$\bar{C}_{fb}^{noisy} = \lim_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy} = \sup_n \bar{C}_{fb,n}^{noisy}$$

for all  $n \geq 1$ . This further implies  $\bar{C}_{fb}^{noisy} \geq \bar{C}_{fb,n}^{noisy}$  for all  $n \geq 1$ .

**Proof.** See Appendix. ■

This lemma indicates that the limit value  $\bar{C}_{fb}^{noisy}$  is lower bounded by any n-block value  $\bar{C}_{fb,n}^{noisy}$ . For a fixed relatively small  $n$ ,  $\bar{C}_{fb,n}^{noisy}$  can be obtained efficiently according to Corollary 54. However, we know that this numerical value is a lower bound on the upper bound of the capacity, which itself is not helpful to evaluate the capacity. In the next lemma, we show that the n-block value  $\bar{C}_{fb,n}^{noisy}$  exponential-block-wise increasingly converges to the limit value  $\bar{C}_{fb}^{noisy}$ . In addition, The convergence trend is exponential-block-wise flatten.

**Lemma 60** *Assume the Gaussian noises  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$  are stationary. Then for any fixed  $n \geq 1$ , we have*

$$\bar{C}_{fb,n}^{noisy} \leq \bar{C}_{fb,kn}^{noisy}$$

for all  $k \geq 1$ . Furthermore, for a fixed  $n$ , let  $\Delta_k = \bar{C}_{fb,2^{k+1}n}^{noisy} - \bar{C}_{fb,2^k n}^{noisy}$ , we have

$$\lim_{k \rightarrow \infty} \Delta_k = 0.$$

Based on the above two lemmas, we conclude that the sequence of  $\{\bar{C}_{fb,n}^{noisy}\}_{n=1}^\infty$  is exponential-block-wise increasing and converges to  $\bar{C}_{fb}^{noisy}$ . However, how to evaluate  $\bar{C}_{fb}^{noisy}$  through  $\bar{C}_{fb,n}^{noisy}$  remains to be seen.

## 5.4 Lower Bound Characterization

In this section, we provide a lower bound  $\underline{C}_{fb,n}^{noisy}$  on the n-block capacity for time-varying Gaussian channels. We show that  $\underline{C}_{fb,n}^{noisy}$  can be obtained by solving a convex optimization problem. Then we consider the stationary Gaussian channels and show that  $\lim_{n \rightarrow \infty} \underline{C}_{fb,n}^{noisy}$  exists and can be represented as a single infinite dimensional optimization problem. According to (5.8), we know  $\lim_{n \rightarrow \infty} \underline{C}_{fb,n}^{noisy}$  is a lower bound on  $C_{fb}^{noisy}$ . However, this limit value is not easy to obtain in general. We next provide two computable lower bounds on this limit value and thus give computable lower bounds on  $C_{fb}^{noisy}$ . Note that, different from the literature of deriving lower bounds (i.e. achievable rates), we herein use a novel approach to find a lower bound instead of proposing a specific coding scheme for the noisy feedback channel, namely, our lower bound is not restricted to any specific coding scheme. The motivation of this novel approach is stated as follows: because linear coding schemes may not achieve any positive transmission rate for noisy feedback channels<sup>2</sup>, propose a specific coding scheme with positive rate may not be an efficient or doable approach to obtain a lower bound on the capacity.

### 5.4.1 Lower Bound On The N-block Capacity

Now, we present a lower bound  $\underline{C}_{fb,n}^{noisy}$  on the n-block capacity  $C_{fb,n}^{noisy}$ .

**Theorem 61** *Consider the noisy feedback Gaussian channels in Fig. 5.1, and a noiseless feedback Gaussian channels with additive noise  $\{W_i + V_i\}_{i=1}^{\infty}$  where noises  $W_i$  and  $V_i$  have the identical statistical properties as that in the noisy feedback settings. See Fig. 5.5 (right) for a vector-representation of this new channel. Denote  $C_{fb,n}^{(w+v)}$  as the n-block capacity of this noiseless feedback Gaussian channel, then*

$$C_{fb,n}^{noisy} \geq C_{fb,n}^{(w+v)}$$

with

$$C_{fb,n}^{(w+v)} = \max_{\mathbf{B}_n, \mathbf{K}_{s,n}} \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n) \mathbf{K}_{wv,n} (\mathbf{I}_n + \mathbf{B}_n)^T + \mathbf{K}_{s,n})}{\det \mathbf{K}_{wv,n}}, \quad (5.21)$$

---

<sup>2</sup>this statement was proved for the case of AWGN channel with AWGN feedback in Kim et al. (2007).

where  $\mathbf{K}_{wv,n} = \mathbf{K}_{v,n} + \mathbf{K}_{w,n}$  and the maximum is taken over all positive semidefinite matrices  $\mathbf{K}_{s,n}$  and all strictly lower triangular matrices  $\mathbf{B}_n$  satisfying

$$\frac{1}{n} \text{tr}(\mathbf{K}_{s,n} + \mathbf{B}_n \mathbf{K}_{wv,n} \mathbf{B}_n') \leq P.$$

In sequel, we denote  $\underline{C}_{fb,n}^{noisy} = C_{fb,n}^{(w+v)}$  as a lower bound on  $C_{fb,n}^{noisy}$ .

**Remark 62** This lower bound is tight when  $\mathbf{K}_{v,n}$  is small (in the positive semi-definite cone) and becomes loose as  $\mathbf{K}_{v,n}$  increases. Note that this lower bound becomes useless when it is below the corresponding  $n$ -block nonfeedback capacity.

**Remark 63** For a fixed  $n$ , the characterization of  $C_{fb,n}^{(w+v)}$  is obtained by

$$C_{fb,n}^{(w+v)} = \frac{1}{n} \max_{\text{admissible coding scheme}} h(Y^n) - h(W^n + V^n)$$

where  $Y^n$  is the channel outputs in Fig. 5.5(right). In addition, it is known that Cover-Pombra scheme achieves this  $n$ -block noiseless feedback capacity. The statements in this remark are justified in Cover and Pombra (1989).

According to Vandenberghe et al. (1998), the above optimization problem can be transformed into the following convex form. The proof is omitted as it directly follows from the proof in Vandenberghe et al. (1998).

**Corollary 64** The lower bound  $\underline{C}_{fb,n}^{noisy}$  on the  $n$ -block noisy feedback capacity can be obtained by solving the following convex optimization problem,

$$\begin{aligned} & \underset{\mathbf{H}_n, \mathbf{B}_n}{\text{maximize}} && \frac{1}{2n} \log \det \mathbf{H}_n - \frac{1}{2n} \log \det \mathbf{K}_{wv,n} \\ & \text{subject to} && \text{tr}(\mathbf{H}_n - \mathbf{K}_{wv,n} \mathbf{B}_n^T - \mathbf{B}_n \mathbf{K}_{wv,n} - \mathbf{K}_{wv,n}) \leq nP, \\ & && \begin{bmatrix} \mathbf{H}_n & \mathbf{I}_n + \mathbf{B}_n^T \\ \mathbf{I}_n + \mathbf{B}_n & \mathbf{K}_{wv,n}^{-1} \end{bmatrix} \geq 0, \\ & && \mathbf{B}_n \text{ is strictly lower triangular.} \end{aligned}$$

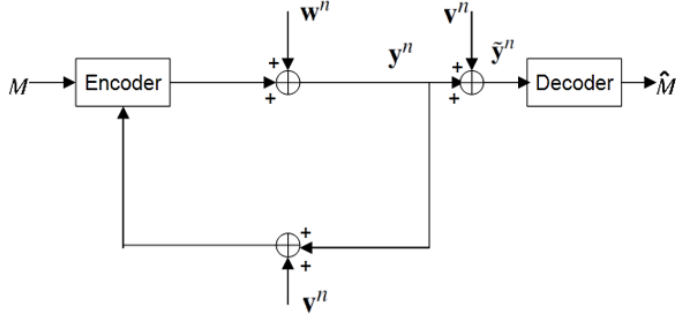


Figure 5.4 New constructed Gaussian channel with noisy feedback

Similarly, the complexity of solving Corollary 64 is evaluated in the following proposition.

**Proposition 65** *The complexity of solving the LMI-optimization problem in Corollary 64 is upper bounded by*

$$\mathcal{O}\left(\frac{81}{8}n^7 + \frac{27}{8}n^6 - \frac{3}{8}n^4 - \frac{1}{8}n^3\right).$$

The proof directly follows from that of Proposition 55 with  $M = 2n + 1$  and  $N = \frac{3}{2}n^2 - \frac{1}{2}n$ . In what follows, we show the proof of Theorem 61. First of all, we need a lemma as follows.

**Lemma 66** *Let  $\mathbf{A} \succeq 0$  and  $\mathbf{C} \succeq \mathbf{B} \succ 0$ , then*

$$\log \frac{\det(\mathbf{A} + \mathbf{B})}{\det \mathbf{B}} \geq \log \frac{\det(\mathbf{A} + \mathbf{C})}{\det \mathbf{C}}$$

**Proof.** See Appendix. ■

Now, we present the proof of Theorem 61.

**Proof.** We first apply the CP-like scheme to the noisy feedback Gaussian channels. Let  $r_n$  be the optimal objective value of the following optimization problem,

$$\begin{aligned} \max_{\mathbf{B}_n, \mathbf{K}_{s,n}} \quad & \frac{1}{2n} \log \frac{\det(\mathbf{I}_n + \mathbf{B}_n) \mathbf{K}_{w,n} (\mathbf{I}_n + \mathbf{B}_n)^T + \mathbf{B}_n \mathbf{K}_{v,n} \mathbf{B}_n^T + \mathbf{K}_{s,n}}{\det(\mathbf{I}_n + \mathbf{B}_n) \mathbf{K}_{w,n} (\mathbf{I}_n + \mathbf{B}_n)^T + \mathbf{B}_n \mathbf{K}_{v,n} \mathbf{B}_n^T} \\ \text{s.t.} \quad & \text{tr}(\mathbf{K}_{s,n} + \mathbf{B}_n(\mathbf{K}_{v,n} + \mathbf{K}_{w,n})\mathbf{B}_n^T) \leq nP, \\ & \mathbf{K}_{s,n} \geq 0, \quad \mathbf{B}_n \text{ is strictly lower triangular.} \end{aligned} \tag{5.22}$$



By invoking the random coding argument used for the achievability proof in Cover and Pombra (1989), we have an achievability result that, under CP-like scheme, there exists a sequence of  $(n, 2^{(r_n - \epsilon)})$  channel codes with  $P_e^n \rightarrow 0$  as  $n \rightarrow \infty$ , for any  $\epsilon$ . The proof of this achievability result is omitted since it directly follows from the achievability proof in Cover and Pombra (1989). Thus, according to the definition of  $C_{fb,n}^{noisy}$ , we clearly have  $C_{fb,n}^{noisy} \geq r_n$ .

We next show  $r_n \geq C_{fb,n}^{(w+v)}$ . Denote  $(\mathbf{B}_n^*, \mathbf{K}_{s,n}^*)$  as an optimal solution of (5.21). Define

$$\mathbf{K}_{\tilde{v},n} = (\mathbf{I}_n - (\mathbf{I}_n + \mathbf{B}_n^*)^{-1})\mathbf{K}_{v,n}(\mathbf{I}_n - (\mathbf{I}_n + \mathbf{B}_n^*)^{-1})^T. \quad (5.23)$$

We clearly have matrix  $\mathbf{K}_v \succeq \mathbf{K}_{\tilde{v}}$ . According to Lemma 66, we have

$$\begin{aligned} C_{fb,n}^{(w+v)} &= \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n^*)(\mathbf{K}_{w,n} + \mathbf{K}_{v,n})(\mathbf{I}_n + \mathbf{B}_n^*)^T + \mathbf{K}_{s,n}^*)}{\det(\mathbf{K}_{w,n} + \mathbf{K}_{v,n})} \\ &= \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n^*)(\mathbf{K}_{w,n} + \mathbf{K}_{v,n})(\mathbf{I}_n + \mathbf{B}_n^*)^T + \mathbf{K}_{s,n}^*)}{\det(\mathbf{I}_n + \mathbf{B}_n^*)(\mathbf{K}_{w,n} + \mathbf{K}_{v,n})(\mathbf{I}_n + \mathbf{B}_n^*)^T} \\ &\leq \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n^*)(\mathbf{K}_{w,n} + \mathbf{K}_{\tilde{v},n})(\mathbf{I}_n + \mathbf{B}_n^*)^T + \mathbf{K}_{s,n}^*)}{\det(\mathbf{I}_n + \mathbf{B}_n^*)(\mathbf{K}_{w,n} + \mathbf{K}_{\tilde{v},n})(\mathbf{I}_n + \mathbf{B}_n^*)^T} \\ &= \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n^*)(\mathbf{K}_{w,n} + \mathbf{K}_{\tilde{v},n})(\mathbf{I}_n + \mathbf{B}_n^*)^T + \mathbf{K}_{s,n}^*)}{\det(\mathbf{K}_{w,n} + \mathbf{K}_{\tilde{v},n})}. \end{aligned}$$

Denote  $r_{w,\tilde{v}}$  as the value of the formula in the last line. We have  $C_{fb,n}^{(w+v)} \leq r_{w,\tilde{v}}$ . By substituting (5.23) and using matrix inverse lemma, we have

$$r_{w,\tilde{v}} = \frac{1}{2n} \log \frac{\det(\mathbf{I}_n + \mathbf{B}_n^*)\mathbf{K}_{w,n}(\mathbf{I}_n + \mathbf{B}_n^*)^T + \mathbf{B}_n^*\mathbf{K}_{v,n}\mathbf{B}_n^{*T} + \mathbf{K}_{s,n}^*}{\det(\mathbf{I}_n + \mathbf{B}_n^*)\mathbf{K}_{w,n}(\mathbf{I}_n + \mathbf{B}_n^*)^T + \mathbf{B}_n^*\mathbf{K}_{v,n}\mathbf{B}_n^{*T}}. \quad (5.24)$$

Thus,  $r_{w,\tilde{v}} \leq r_n$ . Putting above together, we have

$$C_{fb,n}^{noisy} \geq r_n \geq r_{w,\tilde{v}} \geq C_{fb,n}^{(w+v)}$$

The proof is complete. ■

**Remark 67** *Instead of the mathematical proof, we present a heuristic proof of Theorem 61 which provides more insight in this lower bound. First of all, we consider a new channel with noisy feedback as shown in Fig.5.4. An identical Gaussian noise  $V$  is added on the channel output. Since the decoder is not allowed to access the new additive noise, any  $n$ -block achievable*

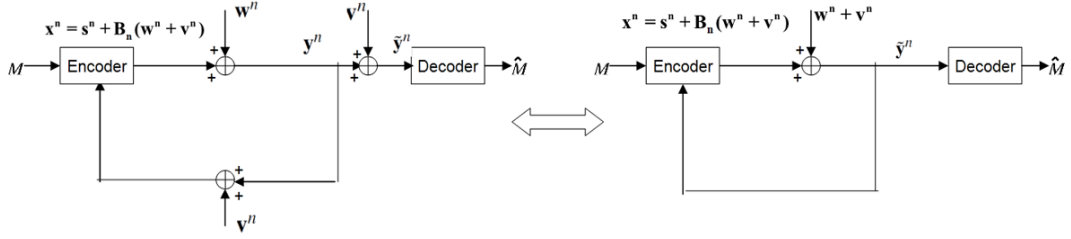


Figure 5.5 The equivalent Gaussian channel with noiseless feedback

rate of the new channel must be a lower bound on the  $n$ -block capacity  $C_{fb,n}^{noisy}$  of the original channel as shown in Fig.5.1. We next apply the CP-like scheme (resulting in an  $n$ -block achievable rate) to this new channel. As a result of the linear feedback scheme, we have the equivalence as shown in Fig.5.5. It is easy to verify the equivalence by checking the channel input and the information received by the decoder. It is known that the optimal CP-like scheme has an achievable rate  $C_{fb,n}^{(w+v)}$ . Thus, we have  $C_{fb,n}^{noisy} \geq C_{fb,n}^{(w+v)}$ .

#### 5.4.2 Characterization under Stationary Gaussian Channels

In this section, we consider stationary Gaussian channels and characterize the limit of the  $n$ -block lower bound which provides a lower bound on the Shannon capacity.

**Theorem 68** Assume that  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$  are stationary processes. Then the limit of the proposed  $n$ -block lower bound (5.21) exists and can be characterized as

$$\underline{C}_{fb}^{noisy} = \sup_{\mathbb{S}_s, \mathbb{B}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \frac{\mathbb{S}_s(e^{i\theta}) + |1 + \mathbb{B}(e^{i\theta})|^2 \mathbb{S}_{wv}(e^{i\theta})}{\mathbb{S}_{wv}(e^{i\theta})} d\theta \quad (5.25)$$

with power constraint

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbb{S}_s(e^{i\theta}) + |\mathbb{B}(e^{i\theta})|^2 \mathbb{S}_{wv}(e^{i\theta}) d\theta \leq P. \quad (5.26)$$

Here,  $\mathbb{S}_w(e^{i\theta})$  and  $\mathbb{S}_v(e^{i\theta})$  are the power spectral density of  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$ , respectively, and  $\mathbb{S}_{wv}(e^{i\theta}) = \mathbb{S}_w(e^{i\theta}) + \mathbb{S}_v(e^{i\theta})$ . The maximization is taken over all power spectral density  $\mathbb{S}_s(e^{i\theta}) \geq 0$  and strictly causal linear filter  $\mathbb{B}(e^{i\theta}) = \sum_{k=1}^\infty b_k e^{ik\theta}$ .

Because the above formula is identical to the capacity formula for the noiseless feedback Gaussian channels with additive Gaussian noise  $\tilde{W}_i = W_i + V_i$ , the proof directly follows from Kim (2010) and thus is omitted. In addition, all the existing solutions of (5.25) on the noiseless feedback capacity (e.g. when  $\tilde{W}_i$  belongs to ARMA(1)) can be used to provide a lower bound on the noisy feedback capacity. However, for arbitrarily stationary Gaussian channels with noiseless feedback, the noiseless feedback capacity is not known yet and thus the lower bound  $\underline{C}_{fb}^{noisy}$  can not be evaluated. In what follows, we provide two approaches to obtain computable lower bounds for arbitrary stationary Gaussian noise. The basic idea is to propose computable lower bounds on  $\underline{C}_{fb}^{noisy}$  instead of evaluating  $\underline{C}_{fb}^{noisy}$ , which is a lower bound on the Shannon capacity.

#### 5.4.2.1 N-block Lower Bound Approach

The first approach invokes the convergence property of super-additive sequence. In the next lemma, we show that, for arbitrarily stationary Gaussian channels, the n-block lower bound  $\underline{C}_{fb,n}^{noisy}$  serves as a lower bound on the capacity  $C_{fb}^{noisy}$ , for any  $n \geq 1$ . We refer readers to Appendix for the introduction on the super-additive sequence and the proof of the lemma.

**Lemma 69** *Assume that the Gaussian noises  $\{W_i\}_{i=1}^{\infty}$  and  $\{V_i\}_{i=1}^{\infty}$  are stationary. Then*

$$\underline{C}_{fb}^{noisy} = \lim_{n \rightarrow \infty} \underline{C}_{fb,n}^{noisy} = \sup_n \underline{C}_{fb,n}^{noisy}$$

*for all  $n \geq 1$ . This further implies  $\underline{C}_{fb}^{noisy} \geq \underline{C}_{fb,n}^{noisy}$  for all  $n \geq 1$ .*

#### 5.4.2.2 LQG Control Approach

According to Theorem 68, we know that any achievable rate (less than or equal to  $\underline{C}_{fb}^{noisy}$ ) of the new noiseless feedback Gaussian channel is a lower bound on the capacity of the original noisy feedback Gaussian channel. Thus, we would like to point out that an alternative lower bound can be computed using control-oriented coding schemes (Elia (2004); Ardestanizadeh and Franceschetti (2012); Ardestanizadeh et al. (2012)) on the new noiseless feedback Gaussian channel. See Fig. 5.6. The basic idea is to design controllers  $K$  that stabilize an unstable,

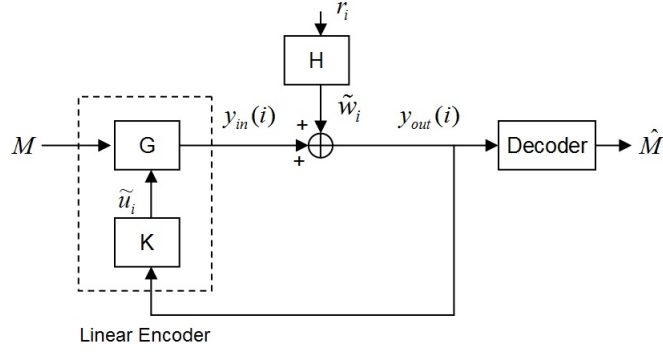


Figure 5.6 LQG coding scheme:  $r \in R^m$  is a vector of white Gaussian noises with zero mean and unit variance without loss of generality, and  $H$  is a stable LTI filter.  $M$  can be treated as an initial condition of the system  $G$ , as described in Elia (2004); Ardestanizadeh and Franceschetti (2012)

single-input-single-output(SISO), linear time-invariant (LTI) system  $G$  in feedback over the given Gaussian channel using the smallest transmission power. For a system  $G$  with given degree of instability, denoted by  $DI$ , the source information can be carried reliably (in the sense of Shannon) through the Gaussian channel at a rate (arbitrarily close) to (at least)  $\log(DI)$ . Note that this rate is independent from the average transmission power. Then the optimal stabilizing controller  $K$  solves a classical LQG regulator problem and provides the smallest transmission power for a given transmission rate. Note that this control-oriented approach is well established in Elia (2004); Ardestanizadeh and Franceschetti (2012); Ardestanizadeh et al. (2012) and a detailed description is outside the scope of this thesis. In the next section we present a numerical example of its application.

## 5.5 Simulation Results

In this section, we present some simulations to verify our results. The forward channel is assumed to be a 1st-order moving average (MV(1)) Gaussian process. Namely,

$$W_i = U_i + \alpha U_{i-1}$$

where  $U_i$  is a white Gaussian process with zero mean and unit variance. Note that a larger  $\alpha$  implies a more correlated channel noise or equivalently a Gaussian channel with stronger

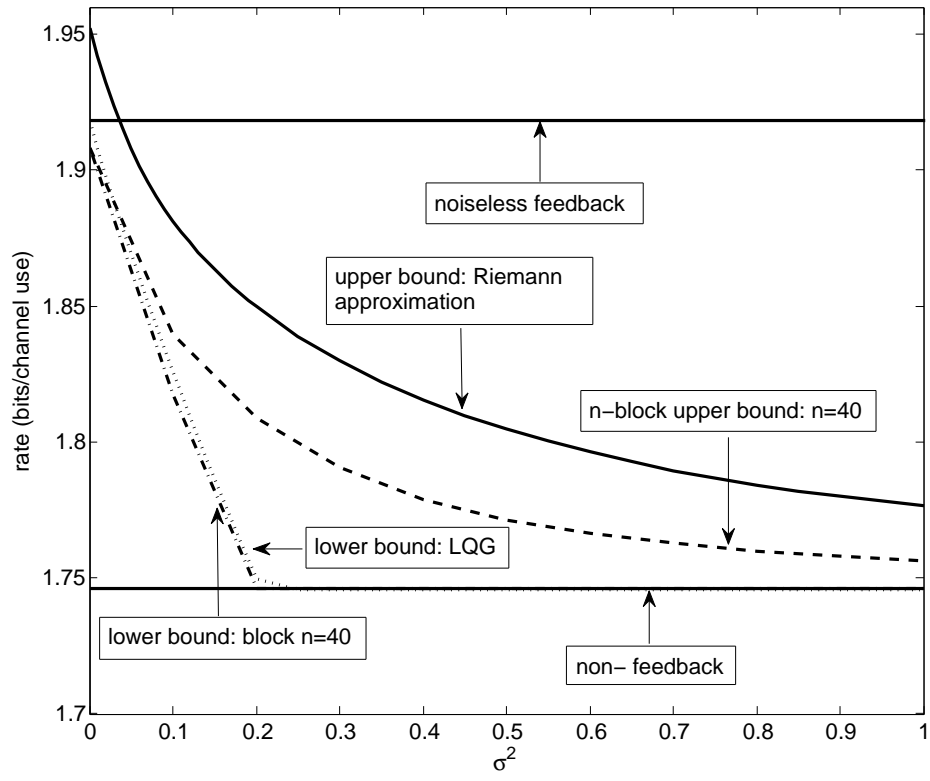


Figure 5.7 Capacity bounds for MV(1) channel with AWGN feedback.

memory. Then

$$\mathbf{K}_{w,n}(i,j) = \begin{cases} 1 + \alpha^2 & i = j \\ \alpha & j = i \pm 1 \\ 0 & \text{else} \end{cases} \quad 1 \leq i, j \leq n.$$

The feedback channel is assumed to be an additive white Gaussian noise with  $\mathbf{K}_{v,n} = \sigma^2 \mathbf{I}_n$ . The average transmission power is  $P = 10$ .

Consider  $\alpha = 0.5$  in our simulations. Fig.5.7 shows the capacity bounds w.r.t. the variance of the feedback noise. The noiseless feedback capacity is obtained by solving a fourth-order polynomial equation in Theorem 5.3 in Kim (2010), and the non-feedback capacity is obtained by solving formula (5.2). For the lower bound on the noisy feedback capacity, we first applied Lemma 69 with  $n = 40$  and find out that the noisy feedback increases the capacity significantly in the regime of small noise power in the feedback. In addition, we obtained another lower bound by solving the LQG control problem as follows. See the new noiseless Gaussian channel as shown in Fig.5.6. The new Gaussian noise  $\tilde{w}_i = w_i + v_i$  can be characterized in the state-space form as

$$\begin{aligned} H : \quad x_{\tilde{w}}[i+1] &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} u[i] \\ v[i] \end{bmatrix} \\ \tilde{w}[i] &= \alpha x_{\tilde{w}}[i] + \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} u[i] \\ v[i] \end{bmatrix} \end{aligned}$$

We now constructed a single state unstable system  $G$  as

$$\begin{aligned} G : \quad x[i+1] &= ax[i] + \tilde{u}_i \\ y_{in}[i] &= x[i]. \end{aligned}$$

where  $a = -2^R$  ( $R$  is a target achievable rate in prior). The lower bound plot is obtained by adjusting  $R$  until the minimized average transmission power  $P^* = 10$ .

For the upper bound, we first show an interesting feature of the n-block upper bound w.r.t. the block length  $n$  at feedback noise variance  $\sigma^2 = 0.3, 0.5, 0.7$ . Fig. 5.8 shows that the n-block upper bound monotonic-increasingly converge to the limit value, and the increase is very gradual for  $n \geq 30$ . To obtain an upper bound (i.e. the limit value of the n-block upper bound) on

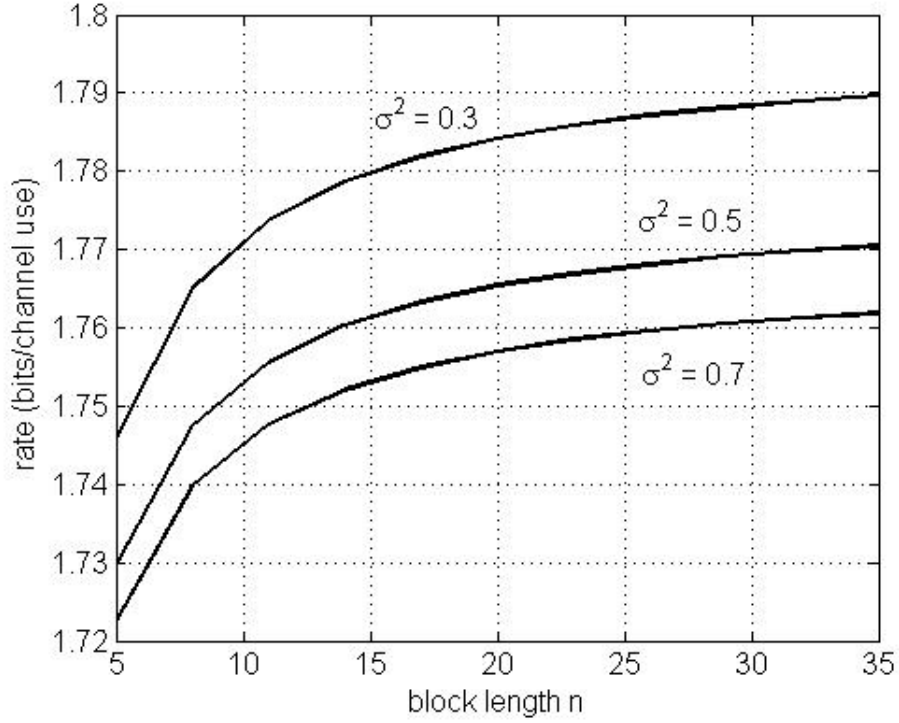


Figure 5.8 Convergence of the  $n$ -block upper bounds ( $\sigma^2 = 0.3, 0.5, 0.7$ ).

the Shannon capacity, we can only approximate it by looking at the  $n$ -block upper bound with sufficiently large  $n$ . Due to the limited practical computing capability, we plotted the  $n$ -block upper bound with  $n = 40$  in Fig. 5.7. Moreover, based on the curves shown in Fig. 5.8, we expect that the curves in Fig. 5.8 will flatten with large  $n$ , while the curve of the limit value of the  $n$ -block upper bound as function of sigma will maintain a similar convex shape as shown in Fig. 5.7. Next, we apply Riemann approximation to obtain an upper bound on the Shannon noisy feedback capacity, which is slightly loose due to the lost of causality constraint on the feedback filter. Combined with the lower bound, we may conclude that the noisy feedback capacity is very sensitive to the feedback noise in MV(1) channel since the upper and lower bound curves are dramatically decreasing as the feedback noise power increases.

## 5.6 Conclusion

In this chapter, we extended the result of Cover-Pombra on the noiseless feedback Gaussian channels to the noisy feedback Gaussian case. We considered the time-varying Gaussian channels with Gaussian noise feedback. We defined the  $n$ -block noisy feedback capacity and derived the lower and upper bounds which can be obtained by solving convex optimization problems. By assuming stationarity on the Gaussian noises, we have characterized the limits of the  $n$ -block upper and lower bounds, which are bounds on the noisy feedback Shannon capacity. We then used Riemann approximation to compute the upper bound and proposed two approaches to obtain computable lower bounds on the noisy feedback capacity. We hope, the results provided in this chapter could foster further advances toward the solution of the noisy feedback capacity problems.



## CHAPTER 6. CAPACITY-ACHIEVING CODES FOR NOISY FEEDBACK CHANNELS: A NECESSARY CONDITION

### 6.1 Introduction

As we have shown in the preceding chapters, by utilizing the information flow decomposition equality, we successfully provide multiple novel results on the capacity of channels with noisy feedback. However, for certain class of channels, using (even noiseless) feedback may not increase the channel capacity, namely, the feedback capacity is equal to the non-feedback capacity which is already known in the literature. Let us name these class of channels as *feedback-unfavorable* channels. For example, discrete memoryless channels (DMCs) are such a well-known class of channels. For channels with memory, we have some examples as follows. In Alajaji (1994) and Alajaji and Fuja (1994), it has been shown that feedback does not increase the capacity of discrete channels with modulo additive noise and channels with memory satisfying the symmetry conditions, respectively. In Shrader and Permuter (2009), it has been shown that for the compound Gilbert-Elliot channel, feedback does not increase capacity. Recently, it is shown that for a class of symmetric finite-state Markov channels, feedback fails to increase capacity (Sen et al. (2011)).

As the capacity for certain feedback-unfavorable channels with noisy feedback is already known, we herein turn our attention to investigate capacity-achieving codes. Searching capacity-achieving feedback code has been a hot topic since the introduction of “feedback communication” and, for the noiseless feedback case, some notable results have been obtained for *feedback-unfavorable* channels, e.g., S-K feedback code for AWGN channels with noiseless feedback (Schalkwijk and Kailath (1966)). However, a careful review finds that all existing capacity-achieving feedback codes are not directly extendable to the noisy feedback settings. For ex-

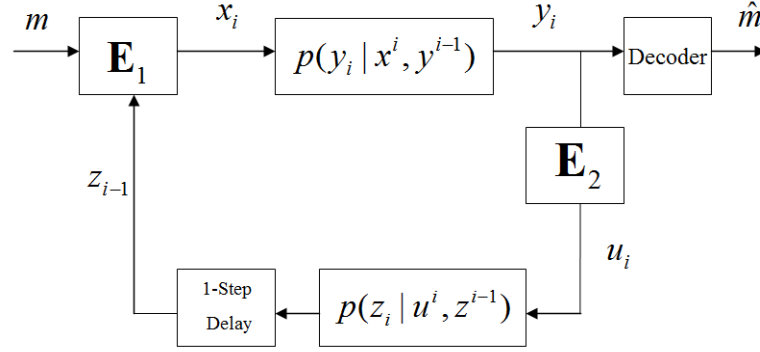


Figure 6.1 Communication channels with noisy feedback: feedback information is allowed to be encoded.

ample, S-K feedback code can not achieve any positive rate for AWGN channels with AWGN feedback (Kim et al. (2007)). The effort of searching capacity-achieving feedback code for noisy feedback channels emerges recently, e.g., Chance and Love (2011b); Martins and Weissman (2008).

In this chapter, we are interested in a high-level problem instead of searching specific capacity-achieving code: is there a capacity-achieving feedback code for *feedback-unfavorable* channels with noisy feedback? If yes, where is it? If no, why? In what follows, by using the information flow decomposition equality, we derive a necessary condition on capacity-achieving channel code, indicating that using noisy feedback is detrimental to the maximal achievable rate (to be precise in the context).

## 6.2 Modeling and Preliminaries

We consider a feedback communication model as shown in Fig.6.1, the channel and the feedback link at time instant  $i$  are modeled as  $p(y_i|x^i, y^{i-1})$  and  $p(z_i|u^i, z^{i-1})$ , respectively. Different from the models used for the proceeding chapters, we now allow another encoder to produce feedback information instead of passively sending back the channel outputs. Specifically, we have message index  $m \in \{1, 2, \dots, 2^{nR}\}$  where  $R$  refers to the transmission rate. At

time instant  $i$ , the encoder  $\mathbb{E}_1$  takes  $m$  and the past feedback information  $z^{i-1}$  to produce the channel input  $x_i = f_i(m, z^{i-1})$  and the encoder  $\mathbb{E}_2$  takes the past channel outputs  $y^i$  to produce feedback input  $u_i = g_i(y^i)$ . After  $n$  time instants, the decoder recovers the message index  $\hat{m}$  by processing channel outputs  $y^n$ . The time ordering of these random variables are presented below<sup>1</sup>.

$$M, X_1, Y_1, U_1, Z_1, \dots, X_n, Y_n, U_n, Z_n, \hat{M}. \quad (6.1)$$

For reader's convenience, we recall the definition of *Probabilistic Limit* and provide some related notations again, as follows.

**Definition 70** (*Probabilistic Limit*) *The limit superior in probability for any sequence  $(X_1, X_2, \dots)$  is defined by*

$$p\text{-}\limsup_{n \rightarrow \infty} X_n = \inf\{\alpha \mid \lim_{n \rightarrow \infty} \text{Prob}\{X_n > \alpha\} = 0\}$$

*Similarly, the limit inferior in probability for any sequence  $(X_1, X_2, \dots)$  is defined by*

$$p\text{-}\liminf_{n \rightarrow \infty} X_n = \sup\{\beta \mid \lim_{n \rightarrow \infty} \text{Prob}\{X_n < \beta\} = 0\}$$

Given a time ordering of random variables  $(X^n, Y^n, Z^n)$  as shown in sequence (6.1),

$$\begin{aligned} \underline{I}(X; Y) &\triangleq p\text{-}\liminf_{n \rightarrow \infty} \frac{1}{n} i(X^n; Y^n) \\ &= p\text{-}\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{p(X^n, Y^n)}{p(X^n)p(Y^n)} \\ \underline{I}(X \rightarrow Y) &\triangleq p\text{-}\liminf_{n \rightarrow \infty} \frac{1}{n} i(X^n \rightarrow Y^n) \\ &= p\text{-}\liminf_{n \rightarrow \infty} \log \frac{\vec{p}_{Y^n|X^n}(Y^n|X^n)}{p_{Y^n}(Y^n)} \end{aligned}$$

where  $\vec{P}_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n p_{Y_i|X^i, Y^{i-1}}(y_i|x^i, y^{i-1})$ .

Now let us recall one useful lemma as follows.

**Lemma 71** (*Han (2003), Verdú and Han (1994)*) *For arbitrary sequence of random variables  $\{X_i\}_{i=0}^\infty$  and  $\{Y_i\}_{i=0}^\infty$ ,*

$$1. \quad p\text{-}\liminf_{n \rightarrow \infty} X_n \leq p\text{-}\limsup_{n \rightarrow \infty} X_n.$$

---

<sup>1</sup>Notice that we do not put any constraint on random variables. Although we will restrict our exposition to finite alphabets in this chapter, all results hold and can be derived in parallel for any abstract set (e.g. countably infinite, continuous alphabets).

2.  $p\text{-}\liminf_{n \rightarrow \infty} (-X_n) = -p\text{-}\limsup_{n \rightarrow \infty} X_n$ .
3.  $p\text{-}\liminf_{n \rightarrow \infty} (X_n + Y_n) \leq p\text{-}\liminf_{n \rightarrow \infty} X_n + p\text{-}\limsup_{n \rightarrow \infty} Y_n$ .
4.  $\underline{I}(X; Y) \leq \liminf_{n \rightarrow \infty} I(X^n, Y^n)$ .

As now we allow the feedback information to be encoded by another encoder at the receiver side, we need re-define the channel code as follows.

**Definition 72** (*Channel Code*) Consider a message  $m$  which is drawn from an index set  $\{1, 2, \dots, 2^{nR}\}$ , a communication channel  $(\mathcal{X}^n, \{p(y_i|x^i, y^{i-1})\}_{i=1}^n, \mathcal{Y}^n)$  with the interpretation that  $X_i$  is the input and  $Y_i$  is the output at time instant  $i$  ( $1 \leq i \leq n$ ) and a feedback link  $(\mathcal{U}^n, \{p(z_i|u^i, z^{i-1})\}_{i=1}^n, \mathcal{Z}^n)$  with the interpretation that  $U_i$  is the input and  $Z_i$  is the output at time instant  $i$ . Then a  $(n, 2^{nR}, \epsilon_n)$  channel code ( $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ ) consists of an index set  $\{1, 2, \dots, 2^{nR}\}$ , a sequence of encoding function  $f_i: \{1, 2, \dots, 2^{nR}\} \times \mathcal{Z}^{i-1} \rightarrow \mathcal{X}_i$ , a sequence of encoding function  $g_i: \mathcal{Y}^i \rightarrow \mathcal{U}_i$ , and a decoding function  $d: \mathcal{Y}^n \rightarrow \{1, 2, \dots, 2^{nR}\}$  with probability of decoding error

$$\frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} p(m \neq d(y^n) | m) \leq \epsilon_n.$$

### 6.3 Rate-Loss in Using Noisy Feedback

In this section, we aim to prove the following theorem.

**Theorem 73** (*A Necessary Condition*) Consider the class of feedback-unfavorable channels with noisy feedback. Then any capacity-achieving channel code must satisfy  $R_L = 0$  where  $R_L \triangleq \underline{I}(Z^{n-1} \rightarrow Y^n)$ .

**Remark 74** As it is shown, the introduced non-negative value  $R_L$  measures the information rate from feedback outputs  $Z$  to channel outputs  $Y$ . Therefore, a channel code with  $R_L = 0$  indicates that the feedback information does not change the distribution of channel output  $Y$ . This further implies that the channel encoder at the transmitter side does not nontrivially/consistently use feedback. Note that if the feedback link is perfect (i.e.  $Z_i = U_i$  for  $\forall i$ ), it

is easy to obtain  $R_L = 0$ . Moreover, if the encoder does not use the feedback or merely uses it for finite time instants, we have  $R_L = 0$  as well.

**Remark 75** *This theorem indicates that it is impossible to find capacity-achieving channel codes by nontrivially using feedback, whereas it is possible in the perfect feedback case (e.g. Schalkwijk-Kailath scheme). Up to present, some feedback coding schemes for feedback-unfavorable channels with noisy feedback have been proposed in the literature. For example, Martins and Weissman (2008) has proposed a linear coding scheme for AWGN channel with bounded feedback noise and Chance and Love (2011b) has proposed a concatenated coding scheme for AWGN channel with noisy feedback. However, these coding schemes cannot achieve the capacity unless, as discussed therein, the feedback additive noise is shrinking to zero (i.e. perfect feedback). Therefore, roughly speaking, using noisy feedback is losing transmission rate. However, it is well known that using (noisy) feedback can improve the reliability (i.e. error exponent) and/or simplify the coding scheme. Thus, we need a tradeoff while using noisy feedback.*

In what follows, we give some technical lemmas before proving the theorem. The first lemma provides information flow decomposition equality (in terms of information density) for the noisy feedback channel as shown in Fig. 6.1.

**Lemma 76** (*Key Lemma*) *For any positive integer  $n$ ,*

$$\begin{aligned} i_{X^n, Y^n}(X^n \rightarrow Y^n) = & i_{M, Y^n}(M, Y^n) + i_{Z^{n-1}, Y^n}(Z^{n-1} \rightarrow Y^n) \\ & + i_{M, Y^n, Z^{n-1}}(M, Z^{n-1} | Y^n) \end{aligned} \tag{6.2}$$

**Proof.** First of all, we have that, for every  $(m, x^n, y^n, z^n, u^n)$  where  $x^i = f^i(m, z^{i-1})$  and  $u^i = g^i(y^i)$ ,

$$\begin{aligned}
& i_{M, Y^n, Z^{n-1}}(m; (y^n, z^{n-1})) \\
&= \log \frac{P_{Y^n, Z^{n-1} | M}(y^n, z^{n-1} | m)}{P_{Y^n, Z^{n-1}}(y^n, z^{n-1})} \\
&= \sum_{i=1}^n \log \frac{P_{Y_i, Z_{i-1} | Y^{i-1}, Z^{i-2}, M}(y_i, z_{i-1} | y^{i-1}, z^{i-2}, m)}{P_{Y_i, Z_{i-1} | Y^{i-1}, Z^{i-2}}(y_i, z_{i-1} | y^{i-1}, z^{i-2})} \\
&= \sum_{i=1}^n \log \left( \frac{P_{Y_i | Y^{i-1}, Z^{i-1}, M}(y_i | y^{i-1}, z^{i-1}, m)}{P_{Y_i | Y^{i-1}, Z^{i-1}}(y_i | y^{i-1}, z^{i-1})} \cdot \frac{P_{Z_{i-1} | Y^{i-1}, Z^{i-2}, M}(z_{i-1} | y^{i-1}, z^{i-2}, m)}{P_{Z_{i-1} | Y^{i-1}, Z^{i-2}}(z_{i-1} | y^{i-1}, z^{i-2})} \right) \\
&\stackrel{(a)}{=} \sum_{i=1}^n \log \frac{P_{Y_i | Y^{i-1}, Z^{i-1}, M}(y_i | y^{i-1}, z^{i-1}, m)}{P_{Y_i | Y^{i-1}, Z^{i-1}}(y_i | y^{i-1}, z^{i-1})} \\
&= \sum_{i=1}^n \log \frac{P_{Y_i | Y^{i-1}, Z^{i-1}, M, X^i}(y_i | y^{i-1}, z^{i-1}, m, x^i)}{P_{Y_i | Y^{i-1}, Z^{i-1}}(y_i | y^{i-1}, z^{i-1})} \\
&\stackrel{(b)}{=} \sum_{i=1}^n \log \frac{P_{Y_i | Y^{i-1}, X^i}(y_i | y^{i-1}, x^i)}{P_{Y_i | Y^{i-1}, Z^{i-1}}(y_i | y^{i-1}, z^{i-1})} \\
&= \log \frac{\vec{P}_{Y^n | X^n}(y^n | x^n)}{P_{Y^n}(y^n)} - \log \frac{\vec{P}_{Y^n | Z^{n-1}}(y^n | z^{n-1})}{P_{Y^n}(y^n)} \\
&= i_{X^n, Y^n}(x^n \rightarrow y^n) - i_{Z^{n-1}, Y^n}(z^{n-1} \rightarrow y^n)
\end{aligned}$$

where (a) and (b) follow from the causality of the channel and the feedback link. By using chain rule,

$$i_{M, Y^n, Z^{n-1}}(m; (y^n, z^{n-1})) = i_{M, Y^n}(m; y^n) + i_{M, Y^n, Z^{n-1}}(m; z^{n-1} | y^n).$$

Putting above equations together,

$$\begin{aligned}
& i_{X^n, Y^n}(x^n \rightarrow y^n) - i_{Z^{n-1}, Y^n}(z^{n-1} \rightarrow y^n) \\
&= i_{M, Y^n}(m; y^n) + i_{M, Y^n, Z^{n-1}}(m; z^{n-1} | y^n).
\end{aligned}$$

The proof is complete. ■

**Lemma 77** For any  $(n, 2^{nR}, \epsilon_n)$  channel code,

$$\underline{I}(M, Z^{n-1} | Y^n) = \liminf_{n \rightarrow \infty} \frac{1}{n} (M, Z^{n-1} | Y^n) = 0.$$

The proof is presented in Appendix. Next, we recall a useful lemma (formula (6) in Alajaji (1994)) as follows.

**Lemma 78** *Every  $(n, 2^{nR}, \epsilon_n)$  channel code satisfies*

$$\epsilon_n \geq \text{Prob}\left\{\frac{1}{n}i(M; Y^n) \leq R - \gamma\right\} - 2^{-\gamma n}$$

for every  $\gamma > 0$ .

Now, we are ready to prove Theorem 73.

**Proof.** Firstly, we show that for every  $(n, 2^{nR}, \epsilon_n)$  channel code

$$R \leq \underline{I}(M; Y^n). \tag{6.3}$$

This can be proved by contradiction as did in Alajaji (1994). Assume that for some  $\rho > 0$ ,

$$R = \underline{I}(M; Y^n) + 2\rho. \tag{6.4}$$

Using Lemma 78 with  $\gamma = \rho$ , we have

$$\epsilon_n \geq \text{Prob}\left\{\frac{1}{n}i(M; Y^n) \leq \underline{I}(M; Y^n) + \rho\right\} - 2^{-\rho n}.$$

Obviously, the righthand term is not vanishing to zero as  $n \rightarrow \infty$  which violates  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ .

Next, according to the inequality (6.3), we have

$$\begin{aligned}
R &\leq \underline{I}(M; Y^n) \\
&= p\text{-}\liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} i_{M, Y^n}(M, Y^n) \right\} \\
&\stackrel{(a)}{=} p\text{-}\liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} i_{X^n, Y^n}(X^n \rightarrow Y^n) - \frac{1}{n} i_{Z^{n-1}, Y^n}(Z^{n-1} \rightarrow Y^n) - \frac{1}{n} i_{M, Y^n, Z^{n-1}}(M, Z^{n-1} | Y^n) \right\} \\
&\stackrel{(b)}{\leq} p\text{-}\liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} i_{X^n, Y^n}(X^n \rightarrow Y^n) \right\} + p\text{-}\limsup_{n \rightarrow \infty} \left\{ -\frac{1}{n} i_{Z^{n-1}, Y^n}(Z^{n-1} \rightarrow Y^n) \right\} \\
&\quad + p\text{-}\limsup_{n \rightarrow \infty} \left\{ -\frac{1}{n} i_{M, Y^n, Z^{n-1}}(M, Z^{n-1} | Y^n) \right\} \\
&\stackrel{(c)}{=} p\text{-}\liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} i_{X^n, Y^n}(X^n \rightarrow Y^n) \right\} - p\text{-}\liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} i_{Z^{n-1}, Y^n}(Z^{n-1} \rightarrow Y^n) \right\} \\
&\quad - p\text{-}\liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} i_{M, Y^n, Z^{n-1}}(M, Z^{n-1} | Y^n) \right\} \\
&\stackrel{(d)}{=} p\text{-}\liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} i_{X^n, Y^n}(X^n \rightarrow Y^n) \right\} - p\text{-}\liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} i_{Z^{n-1}, Y^n}(Z^{n-1} \rightarrow Y^n) \right\} \\
&= \underline{I}(X^n \rightarrow Y^n) - \underline{I}(Z^{n-1} \rightarrow Y^n) \\
&\leq \sup_{\{p(x_i | x^{i-1}, y^{i-1})\}_{i=1}^\infty} \underline{I}(X^n \rightarrow Y^n) - \underline{I}(Z^{n-1} \rightarrow Y^n) \\
&\stackrel{(e)}{=} C_{FB} - \underline{I}(Z^{n-1} \rightarrow Y^n)
\end{aligned}$$

where (a) follows from Lemma 76. Lines (b) and (c) follow from Lemma 71. Line (d) follows from Lemma 77. Line (e) follows from the capacity characterization  $C_{FB}$  in Tatikonda and Mitter (2009) for channels with perfect feedback. Then for feedback-unfavorable channels<sup>2</sup> (i.e.  $C_{FB} = C$ ), we clearly have

$$R \leq C - \underline{I}(Z^{n-1} \rightarrow Y^n). \quad (6.5)$$

Thus, for any capacity-achieving channel code, we must have  $R_L = \underline{I}(Z^{n-1} \rightarrow Y^n) = 0$ . The proof is complete.  $\blacksquare$

**Remark 79**  $R_L$  is nothing but the rate-loss by using noisy feedback. In fact, line (e) holds for general channels and thus may induce generic implications. For example,  $R_L$  should be less

<sup>2</sup>An an illustration Alajaji (1994), for discrete-time finite ( $q$ -ary) alphabet channels with modulo additive noise  $Z$ , we have

$$C_{FB} = C = \log(q) - p\text{-}\limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{P_{Z^n}(Z^n)}.$$



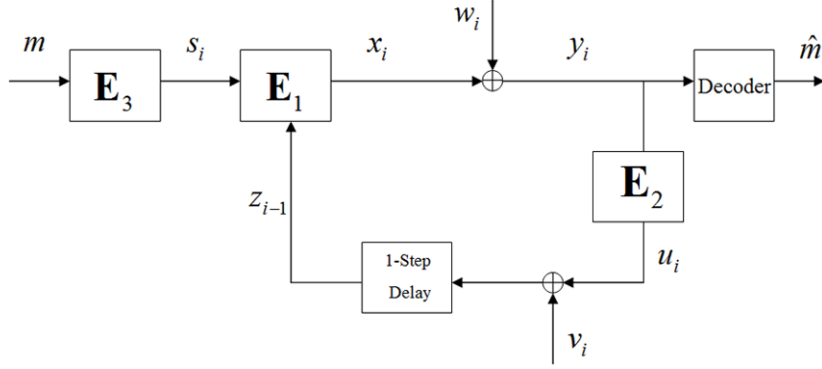


Figure 6.2 A concatenated coding scheme for channels with noisy feedback, where  $\mathbb{E}_3$  represents the outer encoder and  $(\mathbb{E}_1, \mathbb{E}_2)$  represents the inner encoder.

than  $C_{FB} - C$  for noisy feedback to be useful (in the sense of rate). However, we concentrate on the feedback-unfavorable channels since the induced necessary condition has much stronger implication.

**Remark 80** Consider DMCs with noisy feedback and then apply the regular mutual and directed information in the proof of Theorem 73. It is easy to refine the necessary condition to be

$$R_L = \liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} I(Z^{n-1} \rightarrow Y^n) \right\} = 0.$$

## 6.4 An Example: Noisy Feedback AWGN Channels

We now investigate a concatenated coding framework for noisy feedback AWGN channels and show that such a concatenated coding scheme could not achieve the capacity. Consider a concatenated coding scheme<sup>3</sup> as shown in Fig. 6.2, where the inner encoders  $\mathbb{E}_1$  and  $\mathbb{E}_2$  are assumed to be linear time-invariant (LTI) systems and the noises  $W$  and  $V$  are white gaussian noise with zero mean and variance  $\sigma_w^2$  and  $\sigma_v^2$ , respectively. Given a transmission power budget  $P_1$  and  $P_2$  on channel inputs  $X$  and  $U$ , respectively, we know the capacity of this channel is

<sup>3</sup>This coding scheme is studied because it is one possibly implementable feedback coding scheme for AWGN channels with noisy feedback and have been investigated in Chance and Love (2011b).

$$C = \frac{1}{2} \log(1 + \frac{P_1}{\sigma_w}).$$

According to Theorem 73, we now compute rate-loss  $R_L$  for any possible channel codes as follows. Firstly,

$$\begin{aligned} i_{Z^{n-1}, Y^n}(Z^{n-1} \rightarrow Y^n) &= \log \frac{\prod_{i=1}^n P_{Y_i|Z^{i-1}, Y^{i-1}}(Y_i|Z^{i-1}, Y^{i-1})}{P_{Y^n}(Y^n)} \\ &\stackrel{(a)}{=} \log \frac{\prod_{i=1}^n P_{Y_i|V^{i-1}, Y^{i-1}}(Y_i|V^{i-1}, Y^{i-1})}{P_{Y^n}(Y^n)} \\ &= i(V^{n-1} \rightarrow Y^n) \\ &\stackrel{(b)}{=} i(V^{n-1}; Y^n) \end{aligned}$$

where (a) follows from the fact that  $V^{i-1} = Z^{i-1} - U^{i-1}$  and

$$U^{i-1} = g^{i-1}(Y^{i-1}) \triangleq [g_1(Y^1), g_2(Y^2), \dots, g_{i-1}(Y^{i-1})].$$

Line (b) follows from the fact that directed information and mutual information coincide when there exists no feedback from  $Y$  to  $V$  (Massey (1990)). Next, according to Remark 80, we have

$$\begin{aligned} R_L &= \liminf_{n \rightarrow \infty} \frac{1}{n} I(Z^{n-1} \rightarrow Y^n) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} I(V^{n-1}; Y^n) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \{I(V^{n-1}; S^n) + I(V^{n-1}; Y^n|S^n) - I(V^{n-1}; S^n|Y^n)\} \\ &\stackrel{(a)}{=} \liminf_{n \rightarrow \infty} \frac{1}{n} \{I(V^{n-1}; Y^n|S^n) - I(V^{n-1}; S^n|Y^n)\} \\ &\stackrel{(b)}{=} \liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} I(V^{n-1}; Y^n|S^n) \right\} \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \{h(Y^n|S^n) - h(Y^n|S^n, V^{n-1})\} \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \{h((\mathbb{S}W^n)^n + (\mathbb{S}\mathbb{E}_{xz}V^{n-1})^n) - h(\mathbb{S}W^n)^n\} \\ &\quad (\mathbb{S} \text{ and } \mathbb{E}_{xz} \text{ represent the sensitivity function and the transfer function from } Z \text{ to } X, \text{ respectively.}) \\ &\stackrel{(c)}{=} \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( \frac{|\mathbb{S}(e^{j2\pi\theta})|^2(\sigma_w^2 + |\mathbb{E}_{xz}(e^{j2\pi\theta})|^2\sigma_v^2)}{|\mathbb{S}(e^{j2\pi\theta})|^2\sigma_w^2} \right) d\theta \\ &= \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( 1 + |\mathbb{E}_{xz}(e^{j2\pi\theta})|^2 \frac{\sigma_v^2}{\sigma_w^2} \right) d\theta \end{aligned}$$

where (a) follows from the fact that  $V$  is independent of  $S$ . Line (b) can be verified by Fano's inequality. Line (c) follows from the fact that  $V$  and  $W$  are assumed to be white gaussian and then the Szegö-Kolmogorov-Krein theorem applies.

If there exists a capacity-achieving channel code, we must have  $R_L = 0$  which implies  $\mathbb{E}_{xz}(e^{j2\pi\theta}) = 0$  at any frequency (i.e. feedback is not used in the sense of frequency domain). Further the rate-loss  $R_L$  is independent from encoders  $\mathbb{E}_2$  and  $\mathbb{E}_3$  and clearly  $R_L = 0$  if  $\sigma_v^2 = 0$  (i.e. perfect feedback).

## 6.5 Conclusion

In this chapter, we consider a class of feedback-unfavorable channels with noisy feedback, for which the capacity equals to the known non-feedback capacity. As the capacity is known at this point, we turned out attention to find conditions on capacity-achieving channel codes. By investigating the rate loss of using noisy feedback, we have derived a necessary condition on the capacity-achieving channel codes. This condition delivers a negative message that using noisy feedback could not induce any capacity-achieving channel code, or equivalently, is detrimental to the maximal achievable rate.

## CHAPTER 7. NOISY FEEDBACK COMMUNICATIONS WITH SIDE INFORMATION AT THE DECODER

### 7.1 Introduction

In this chapter, we consider noisy feedback channels with side information at the decoder as shown in Fig.7.1 (Li and Elia (2012)). Different from the classical noisy feedback channels considered in preceding chapters, it is assumed that the receiver has access to the information received by the transmitter with finite delays. With this side information, the receiver causally knows what the transmitter will do while the transmitter does not know the full information received by the receiver. In other words, the feedback is noiseless on the receiver side while it is noisy on the transmitter side. Investigating this new framework will bridge the gap between the perfect feedback and the classical noisy feedback and is a key step forward to solve the classical noisy feedback problem.

This framework is motivated by many practical applications of interacting networked systems. One practical example is force feedback for virtual reality applications as shown in Fig. 7.2. The controller/encoder sends force commands to a remote actuator that acts on an object, say squeezing a rubber ball to fix the idea. The commands are delivered through a noisy channel and translated into actuation on the object. Load cells directly measure the actual force exerted on the object but with sensor/measuring noise. The noisy signals from the load cells are fed back to the transmitter over a noiseless channel to provide force feedback to the operator/controller, which can then adjust the grasp strength. Note that, although the actual communication channel in the feedback is noiseless, the effective feedback communication is noisy due to the sensor noise. The load cell measurers can also be used locally at the receiver side to improve the fidelity of the reproduced forces. For this example, we are interested

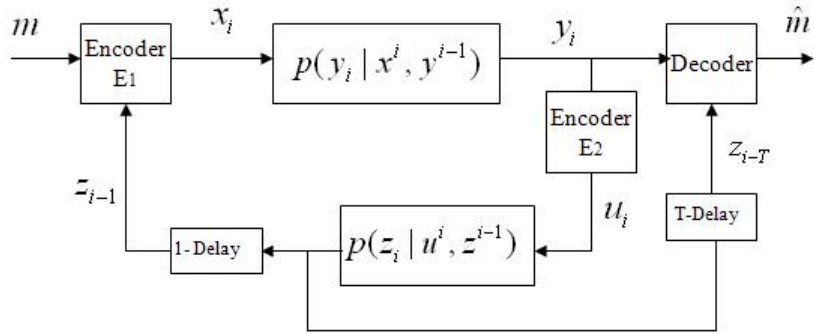


Figure 7.1 Noisy feedback communication channels with side information at the decoder: the feedback information  $Z_i$  is known by the decoder after  $T$ -step delay.

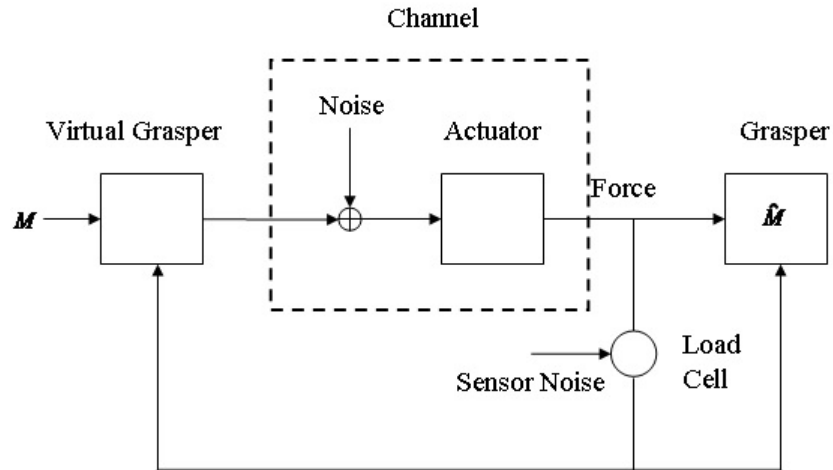


Figure 7.2 Grasper with noisy force feedback

in a question, saying, what is the maximal transmission rate of this noisy actuation channel?

This new framework has many merits and only the most interesting two of them are investigated in this chapter. Firstly, the capacity of noisy feedback channels under our new framework can be characterized by the causal conditional directed information, which is automatically an upper bound on the capacity of the classical noisy feedback channels. Secondly, for certain channels, the new framework allows linear coding schemes with significant positive rate, which can not be easily obtained for the classical noisy feedback framework<sup>1</sup>.

## 7.2 Modeling

As shown in Fig. 7.1, we model the forward channel and the feedback channel at time instant  $i$  as  $p(y_i|x^i, y^{i-1})$  and  $p(z_i|u^i, z^{i-1})$ , respectively. The message index is  $m \in \{1, 2, \dots, 2^{nR}\}$  where  $R$  refers to the transmission rate. At time instant  $i$ , the transmitter  $\mathbb{E}_1$  takes  $m$  and past feedback information  $z^{i-1}$  to produce channel input  $x_i(m, z^{i-1})$  and the feedback transmitter  $\mathbb{E}_2$  takes all past channel outputs  $y^i$  to produce feedback signal  $u_i$ . Different from the classical noisy feedback framework, the receiver(decoder) is allowed to have access to the feedback output  $z_i$  with finite delay  $T$ . After  $n$  time instants, the decoder recovers the message index  $\hat{m}(y^n, z^{n-T})$ . The time ordering of these random variables are presented below.

$$M, X_1, Y_1, U_1, Z_1, \dots, X_{n-1}, Y_{n-1}, U_{n-1}, Z_{n-1}, X_n, Y_n, U_n, Z_n, \hat{M}. \quad (7.1)$$

Next, for reader's convenience, we recall the channel causality of communication channels as shown in Fig. 7.1.

**Definition 81** (*Channel Causality*) Consider random variables  $(M, X^n, Y^n, Z^n, \hat{M})$  in time ordering (7.1). A communication channel  $p(Y_i|X^i, Y^{i-1})$  is causal if

$$p(Y_i|X^i, Y^{i-1}, U^{i-1}, Z^{i-1}, M) = p(Y_i|X^i, Y^{i-1}).$$

Similarly, a feedback channel  $p(Z_i|U^i, Z^{i-1})$  is causal if

$$p(Z_i|U^i, Z^{i-1}, X^i, Y^i, M) = p(Z_i|U^i, Z^{i-1}).$$

---

<sup>1</sup>In Kim et al. (2007), the author proved that for the noisy feedback gaussian channels, linear coding scheme could not achieve any positive transmission rate.

Following the literature, we assume the channel causality in this chapter as well.

### 7.3 Capacity Characterization

For channels without feedback, the capacity is well characterized by the mutual information. For channels with perfect feedback, the capacity can be characterized by the directed information. In what follows, we show that, for noisy feedback channels under the framework as shown in Fig.7.1, the capacity can be characterized by the causal conditioning directed information. The following theorems are of importance to be well understood.

**Theorem 82** (*Main Theorem*) *For noisy feedback channels with side information at the decoder, as shown in Fig.7.1,*

$$\begin{aligned} I(M; (Y^n, Z^{n-T})) \\ = I(X^n \rightarrow Y^n || Z^{n-1}) - I(M; Z_{n-T+1}^{n-1} | Y^n, Z^{n-T}). \end{aligned} \quad (7.2)$$

**Proof.** Firstly, we have

$$\begin{aligned} & i(M; (Y^n, Z^{n-1})) \\ &= \log \frac{p(Y^n, Z^{n-1} | M)}{p(Y^n, Z^{n-1})} \\ &= \sum_{i=1}^n \log \frac{p(Y_i, Z_{i-1} | Y^{i-1}, Z^{i-2}, M)}{p(Y_i, Z_{i-1} | Y^{i-1}, Z^{i-2})} \\ &= \sum_{i=1}^n \log \frac{p(Y_i | Y^{i-1}, Z^{i-1}, M) p(Z_{i-1} | Y^{i-1}, Z^{i-2}, M)}{p(Y_i | Y^{i-1}, Z^{i-1}) p(Z_{i-1} | Y^{i-1}, Z^{i-2})} \\ &= \sum_{i=1}^n \log \frac{p(Y_i | Y^{i-1}, Z^{i-1}, M) p(Z_{i-1} | U^{i-1}, Y^{i-1}, Z^{i-2}, M)}{p(Y_i | Y^{i-1}, Z^{i-1}) p(Z_{i-1} | U^{i-1}, Y^{i-1}, Z^{i-2})} \\ &\stackrel{(a)}{=} \sum_{i=1}^n \log \frac{p(Y_i | Y^{i-1}, Z^{i-1}, M)}{p(Y_i | Y^{i-1}, Z^{i-1})} \\ &= \sum_{i=1}^n \log \frac{p(Y_i | X^i, Y^{i-1}, Z^{i-1}, M)}{p(Y_i | Y^{i-1}, Z^{i-1})} \\ &\stackrel{(b)}{=} \sum_{i=1}^n \log \frac{p(Y_i | X^i, Y^{i-1}, Z^{i-1})}{p(Y_i | Y^{i-1}, Z^{i-1})} \\ &= i(X^n \rightarrow Y^n || Z^{n-1}) \end{aligned}$$

where (a) and (b) follow from the channel causality assumption. Then, by the chain rule of the mutual information, we have

$$\begin{aligned}
& I(M; (Y^n, Z^{n-T})) \\
&= I(M; (Y^n, Z^{n-1})) - I(M; Z_{n-T+1}^{n-1} | Y^n, Z^{n-T}) \\
&= I(X^n \rightarrow Y^n || Z^{n-1}) - I(M; Z_{n-T+1}^{n-1} | Y^n, Z^{n-T}).
\end{aligned}$$

The proof is complete. ■

**Remark 83** We are looking at the mutual information between the message  $M$  and the information  $(Y^n, Z^{n-T})$  because  $(Y^n, Z^{n-T})$  are the actual information received by the decoder, which can be used to recover the message.

**Remark 84** Since the sequence  $Z_{n-T+1}^{n-1}$  in the second term has fixed length  $T$ , the quantity  $I(M; Z_{n-T+1}^{n-1} | Y^n, Z^{n-T})$  should be uniformly bounded (and zero if  $T = 0, 1$ ). If we average both sides in (7.2) over  $n$  and take  $n \rightarrow \infty$ , the bounded quantity  $I(M; Z_{n-T+1}^{n-1} | Y^n, Z^{n-T})$  will shrink to zero. Thus the causal conditional directed information  $I(X^n \rightarrow Y^n || Z^{n-1})$  is the only relevant quantity to characterize the capacity. Without loss of generality, we only consider the case  $T = 1$  (i.e.  $I(M; Z_{n-T+1}^{n-1} | Y^n, Z^{n-1}) = 0$ ) in the rest of this section.

Now we wish to realize the merit of Theorem 82. First of all, we give a high-level discussion on the development of the capacity characterization for perfect feedback channels. Massey (1990) introduced directed information and then it is found that

$$I(M; Y^n) = I(X^n \rightarrow Y^n). \quad (7.3)$$

In light of this equality, it has been shown that the directed information can be used for characterizing the perfect feedback capacity. Although different approaches have been adopted for proving the capacity, the key idea behind all of them is to apply the approaches used in the non-feedback case to the perfect feedback case, based on equation (7.3).

Specifically, there are three avenues for proving the nonfeedback capacity. The first one is attributed to Shannon of using asymptotic equipartition property (AEP), joint typicality decoding. Along this avenue, based on equation (7.3), Cover and Pombra (1989) characterized



the capacity of Gaussian channel with perfect feedback and Kim (2010) presented the capacity of a class of stationary perfect feedback channels. The second one is attributed to Gallager of investigating random coding exponent which was later applied in characterizing the finite-state channel capacity (Gallager (1968)). Along this avenue, based on equation (7.3), Permuter et al. (2009) characterized the capacity of channels with time-invariant deterministic feedback. The third one is attributed to Feinstein's lemma which was applied in characterizing the general channel capacity by Verdú and Han (1994). Then, based on equation (7.3), Tatikonda Tatikonda and Mitter (2009) proved the general perfect feedback capacity accordingly. In a word, equation (7.3) has played a key role in proving the perfect feedback capacity.

In what follows, we take into account the Gaussian noisy feedback channels with side information at the decoder<sup>2</sup>. See Fig.7.3. As it will be shown, based on Theorem 82, we can explicitly characterize its n-block capacity along the first avenue. We believe that the capacity for other classes of noisy feedback channels under our new framework can be also characterized by the causal conditional directed information and the aforementioned three main avenues are adoptable.

### 7.3.1 Capacity of Gaussian Noisy Feedback Channels with Side Information at the Decoder

We provide the capacity characterization in the following theorem and present the proof in Appendix, where Theorem 82 plays a key role in the proof.

**Theorem 85** *Consider Gaussian noisy feedback channels with side information at the decoder,*

---

<sup>2</sup>If the power constraint  $P_2$  is removed and the encoder  $\mathbb{E}_2$  is a time-invariant deterministic gain, this framework essentially converges to the framework considered in Permuter et al. (2009).

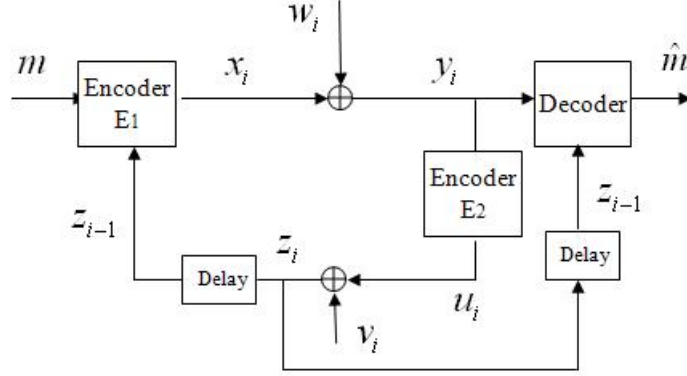


Figure 7.3 Additive Gaussian channel with additive Gaussian noise feedback:  $W^n \sim \mathbb{N}_n(0, \mathbf{K}_{w,n})$  and  $V^n \sim \mathbb{N}_n(0, \mathbf{K}_{v,n})$  where  $\mathbf{K}_{w,n}$  and  $\mathbf{K}_{v,n}$  are non-singular covariance matrices. Average power constraints  $P_1$  and  $P_2$  are implemented on inputs  $X$  and  $U$ , respectively.

as shown in Fig. 7.3. The  $n$ -block capacity<sup>3</sup> is

$$C_{FB,n} = \max_{\mathbf{B}_n, \mathbf{D}_n, \mathbf{K}_{s,n}} \frac{1}{2n} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n) \mathbf{K}_{w,n} (\mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n)^T + \mathbf{K}_{s,n})}{\det \mathbf{K}_{w,n}}$$

$$s.t. \quad \text{tr}(\mathbf{K}_{s,n} + \mathbf{B}_n \mathbf{K}_{v,n} \mathbf{B}_n^T + \mathbf{B}_n \mathbf{D}_n \mathbf{K}_{w,n} \mathbf{D}_n^T \mathbf{B}_n^T) \leq nP_1,$$

$$\text{tr}(\mathbf{D}_n \mathbf{K}_{s,n} \mathbf{D}_n^T + \mathbf{D}_n (\mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n) \mathbf{K}_{w,n} (\mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n)^T \mathbf{D}_n^T + \mathbf{D}_n \mathbf{B}_n \mathbf{K}_{v,n} \mathbf{B}_n^T \mathbf{D}_n^T) \leq nP_2,$$

$$\mathbf{K}_{s,n} \geq 0, \quad \mathbf{B}_n \text{ is strictly lower triangular,}$$

$$\mathbf{D}_n \text{ is lower triangular.}$$

(7.4)

**Remark 86** An interpretation of matrices  $\mathbf{B}_n$  and  $\mathbf{D}_n$  as a specific coding scheme is shown in Fig. E.1 in Appendix. Some notes on the above capacity characterization are presented below,

1. If  $\mathbf{D}_n = \mathbf{0}_n$ , the  $n$ -block capacity formula (7.4) converges to the non-feedback capacity characterization in Cover and Pombra (1989).

<sup>3</sup>As stated in Cover and Pombra (1989), the  $n$ -block capacity  $C_{FB,n}$  is defined as follows. For any  $\epsilon > 0$ , there exists a sequence of  $(n, 2^{n(C_{FB,n}-\epsilon)})$  channel codes with  $P_e^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$ . Conversely, for any  $\epsilon > 0$ , any sequence of  $(n, 2^{n(C_{FB,n}+\epsilon)})$  channel codes has  $P_e^{(n)}$  bounded away from zero for all  $n$ .

2. If the power constraint  $P_2$  is removed and the channel outputs  $Y$  is fed back without any encoding (i.e.  $\mathbf{D}_n = \mathbf{I}_n$ ), the optimization problem (7.4) converges to the upper bound of the classical noisy feedback Gaussian channels, characterized in Li and Elia (2011d).
3. According to the first power constraint  $P_1$ , if the additive noise  $V^n$  is large (in the sense of covariance matrix), the matrix  $\mathbf{B}_n$  will shrink to a zero matrix which implies the feedback turns to be "shut-off".

### 7.3.2 Calculation of A Lower Bound: An Iteration Algorithm

In this section, we propose an iteration algorithm to compute  $C_{FB,n}$ . Note, however, that this algorithm only guarantees local optimality and the global optimality proof is left to our future research. We first present a necessary corollary as follows, with the proof in Appendix.

**Corollary 87** *The optimization problem (85) can be casted into the following form,*

$$\begin{aligned}
 & \underset{\mathbf{Y}_n, \mathbf{B}_n, \mathbf{D}_n, \mathbf{H}_n, \mathbf{Z}_n}{\text{maximize}} && \frac{1}{2n} \log \frac{\det \mathbf{Y}_n}{\det \mathbf{K}_{w,n}} \\
 & s.t. && \begin{bmatrix} \mathbf{H}_n + \mathbf{K}_{w,n} \mathbf{D}_n^T \mathbf{B}_n^T + \mathbf{B}_n \mathbf{D}_n \mathbf{K}_{w,n} + \mathbf{K}_{w,n} - \mathbf{Y}_n & \mathbf{B}_n \\ * & \mathbf{K}_{v,n}^{-1} \end{bmatrix} \geq 0, \\
 & && \begin{bmatrix} \mathbf{Z}_n - \mathbf{D}_n \mathbf{Y}_n \mathbf{D}_n^T & \mathbf{D}_n \mathbf{B}_n \\ * & \mathbf{K}_{v,n}^{-1} \end{bmatrix} \geq 0, \\
 & && \begin{bmatrix} \mathbf{Y}_n & \mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n \\ * & \mathbf{K}_{w,n}^{-1} \end{bmatrix} \geq 0 \\
 & && \text{tr}(\mathbf{H}_n) \leq nP_1, \quad \text{tr}(\mathbf{Z}_n) \leq nP_2, \\
 & && \mathbf{B}_n \text{ is strictly lower triangular,} \\
 & && \mathbf{D}_n \text{ is lower triangular.}
 \end{aligned} \tag{7.5}$$

As it is shown, the constraints in (7.5) are subject to coupling terms  $\mathbf{D}_n \mathbf{Y}_n \mathbf{D}_n^T$ ,  $\mathbf{D}_n \mathbf{B}_n$  and  $\mathbf{B}_n \mathbf{D}_n$ , which are not easy to decouple or substitute. The idea of the iteration algorithm is the

following. With fixed feedback encoder  $\mathbf{D}_n$ , the problem (7.5) turns to be convex and thus can be solved efficiently by the semidefinite programming. So we wish to find a good  $\mathbf{D}_n$  to obtain a larger rate. We herein take the encoder  $\mathbf{D}_n$  with the minimal power cost subject to the fixed  $\mathbf{Y}_n$  and  $\mathbf{B}_n$ . In other words, for a given transmission rate, we wish to design an encoder  $\mathbf{D}_n$  such that the total power cost is minimized. This task can be accomplished through solving the following convex optimization problem.

$$\begin{aligned}
& \underset{\mathbf{D}_n}{\text{minimize}} \quad \tau_1 + \tau_2 \\
& \text{tr}(\mathbf{Y}_n - \mathbf{K}_{w,n}\mathbf{D}_n\mathbf{B}_n^T - \mathbf{B}_n\mathbf{D}_n\mathbf{K}_{w,n} - \mathbf{K}_{w,n} + \mathbf{B}_n\mathbf{K}_{v,n}\mathbf{B}_n) \leq \tau_1, \\
& \text{tr}(\mathbf{D}_n\mathbf{Y}_n\mathbf{D}_n^T + \mathbf{D}_n\mathbf{B}_n\mathbf{K}_{v,n}\mathbf{B}_n^T\mathbf{D}_n^T) \leq \tau_2, \\
& \begin{bmatrix} \mathbf{Y}_n & \mathbf{I}_n + \mathbf{B}_n\mathbf{D}_n \\ * & \mathbf{K}_{w,n}^{-1} \end{bmatrix} \geq 0, \\
& \mathbf{D}_n \text{ is lower triangular, } \quad \tau_1 \leq nP_1, \quad \tau_2 \leq nP_2.
\end{aligned}$$

By introducing dummy matrix  $\mathbf{G}_n$  and invoking Schur complement decomposition, we obtain a convex form below.

$$\begin{aligned}
& \underset{\mathbf{Y}_n, \mathbf{B}_n, \mathbf{D}_n, \mathbf{G}_n}{\text{minimize}} \quad \tau_1 + \tau_2 \\
& \text{s.t.} \quad \begin{bmatrix} \mathbf{Y}_n & \mathbf{I}_n + \mathbf{B}_n\mathbf{D}_n \\ * & \mathbf{K}_{w,n}^{-1} \end{bmatrix} \geq 0, \quad \begin{bmatrix} \mathbf{G}_n & \mathbf{D}_n \\ * & (\mathbf{Y}_n + \mathbf{B}_n\mathbf{K}_{v,n}\mathbf{B}_n^T)^{-1} \end{bmatrix} \geq 0, \\
& \text{tr}(\mathbf{Y}_n - \mathbf{K}_{w,n}\mathbf{D}_n\mathbf{B}_n^T - \mathbf{B}_n\mathbf{D}_n\mathbf{K}_{w,n} - \mathbf{K}_{w,n} + \mathbf{B}_n\mathbf{K}_{v,n}\mathbf{B}_n) \leq \tau_1, \\
& \text{tr}(\mathbf{G}_n) \leq \tau_2, \quad \mathbf{D}_n \text{ is lower triangular,} \\
& \tau_1 \leq nP_1, \quad \tau_2 \leq nP_2.
\end{aligned} \tag{7.6}$$

We are now ready to present the iteration algorithm.

**Algorithm 88** (*Iteration Algorithm*)

1. *Initialization: Set  $n = 0$  (the iteration number), the maximum number of iterations  $n_{max}$ , the desired accuracy  $\delta > 0$ , and the initial feedback encoder  $\mathbf{D}_n$ .*  
*repeat*
2.  $n \leftarrow n + 1$ ;
3. *For fixed current  $\mathbf{D}_n$ , solve the optimization problem (7.5) and obtain the cost value  $t_n$ , matrix  $\mathbf{Y}_n$  and  $\mathbf{B}_n$ .*
4. *For fixed current  $\mathbf{Y}_n$  and  $\mathbf{B}_n$ , solve the optimization problem (7.6) and obtain  $\mathbf{D}_n$ .*  
*until  $n \geq n_{max}$ , or  $t_n - t_{n-1} \leq \delta$ .*

**Theorem 89** *If the optimization problem (7.5) is feasible for the initial feedback encoder  $\mathbf{D}_n$ , the convergence of the sequence of objective value  $t$ 's generated by the algorithm is guaranteed.*

**Proof.** Let  $\mathbf{D}_n$  be a feasible encoder for the problem (7.5) and let objective value  $t_k$  (achievable rate at iteration  $k$ ),  $\mathbf{Y}_n$  and  $\mathbf{B}_n$  be the solution obtained by solving problem (7.5) for the fixed  $\mathbf{D}_n$ . Then for the fixed matrix  $\mathbf{Y}_n$  (i.e. fixed achievable rate) and  $\mathbf{B}_n$ , solve the problem (7.6) will yield a less total power cost (i.e.  $\tau_1 + \tau_2 \leq nP_1 + nP_2$ ) since at least the encoder  $\mathbf{D}_n$  is a feasible solution. Then fix  $\mathbf{D}_n$  again, solve the problem (7.5) will obviously provide a bigger  $t_{k+1}$  (i.e. higher achievable rate). Therefore, the proposed iterative procedure yields a non-decreasing sequence of objective value  $t$ 's which is clearly bounded above. The proof is complete. ■

We end this section by showing some numerical examples.

Fig.7.4 shows the achievable rate calculated by the iteration algorithm, compared with the perfect feedback capacity and the non-feedback capacity. These plots indicate that our new framework can achieve much larger transmission rate (in small feedback noise region) than the non-feedback capacity.

To verify the benefits of using the encoder  $\mathbb{E}_2$ , we now compare the achievable rates obtained by using  $\mathbb{E}_2$  and unit-gain (i.e.  $\mathbb{E}_2 = \mathbf{I}_n$ ) feedback, respectively. Consider the same channel model described in Fig.7.4 with  $P_1 = 10$  and  $\sigma^2 = 0.1$ . For unit-gain feedback, the feedback

transmission power  $P_2 = 12.17$  and the achievable rate  $R = 1.7562$ . However, by applying  $\mathbb{E}_2$  under the same power constraint  $P_2$ , the achievable rate is  $R = 1.7577$ . Note that this rate enhancement may be much more significant for other settings.

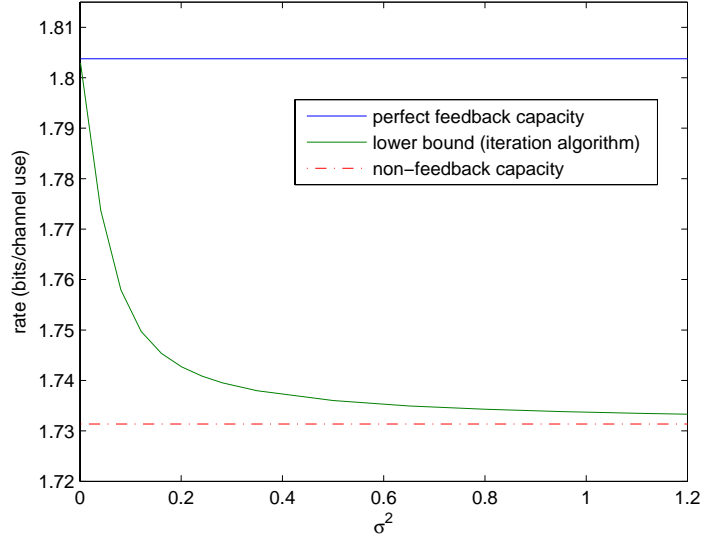


Figure 7.4 The  $n$ -block achievable rate of a Gaussian noisy feedback channel with side information at the decoder: The forward channel is assumed to be a first-order moving average (1st-MV) Gaussian process, Namely,  $W_i = \tilde{U}_i + 0.2\tilde{U}_{i-1}$  where  $\tilde{U}_i$  is a white Gaussian process with zero mean and unit variance; The feedback channel is assumed to be an additive white Gaussian noise with variance  $\sigma^2$ ; The power constraint is  $P_1 = P_2 = 10$  and the coding block length  $n = 30$ .

## 7.4 Simple Coding Strategy

As it is shown in Kim et al. (2007), any linear coding scheme fails to achieve positive transmission rate for the classical Gaussian noisy feedback channels. In contrast, in this section, we aim to show that for certain noisy feedback channels, the new framework allows us to design linear coding scheme to communicate with significant positive transmission rate<sup>4</sup>. In what follows, we investigate a first-order moving average Gaussian channel with random intermittent

<sup>4</sup>A linear coding strategy for AWGN channels with bounded noise feedback which actually fits in our new framework can be found in Martins and Weissman (2008).

feedback (Fig. 7.5). Specifically,

$$W_i = \tilde{U}_i + \alpha \tilde{U}_{i-1}$$

where  $\tilde{U}_i$  ( $i = 0, 1, 2, \dots, n$ ) is an independent identical distributed normal distribution with zero mean and unit variance. The power constraints on channel inputs  $\{X_i\}_{i=0}^\infty$  and feedback inputs  $\{U_i\}_{i=0}^\infty$  are assumed to be  $P_1$  and  $P_2$ , respectively. The feedback information is independently either erased (i.e.  $\xi_i = 0$ ) with probability  $e$  or perfectly received by the transmitter (i.e.  $\xi_i = 1$ ) with probability  $1 - e$ . Now, we present a specific linear coding scheme whose achievable rate is not only positive but also, for some cases, larger than the non-feedback capacity.

Consider a simple coding strategy, denoted by symbol  $\mathcal{C}$ , as follows. Let  $r$  be the real root with the maximal absolute value of the third-order polynomial

$$(1 + \alpha^2)x^3 + 2\alpha x^2 - (1 + \alpha^2 + P_1)x - 2\alpha = 0 \quad (7.7)$$

We first set the encoder  $\mathbb{E}_2$  to be a positive gain

$$g \leq \sqrt{\frac{P_2}{P_1 + 1 + \alpha^2}}.$$

Note that, without reducing the achievable rate of our proposed transmission scheme, the gain  $g$  can be time-varying if it is practically necessary. Now, we have  $U_i = gY_i$ . Next, for encoder  $\mathbb{E}_1$ , define

$$\begin{aligned} A(\xi_i = 0) &= 1, & B(\xi_i = 0) &= 0 \\ A(\xi_i = 1) &= -r, & B(\xi_i = 1) &= r - \frac{1}{r} \end{aligned}$$

At time 0, the channel input is  $X_0$  (the message to transmit) and the decoder receives  $Y_0$ . At time 1, the channel input is

$$X_1 = A(\xi_0)X_0 + \frac{1}{g}B(\xi_0)Z_0.$$

Note that the one-step delay is captured here and  $\xi_i$  is available to the transmitter through detecting  $Z_{i-1} = \{U_{i-1}, \emptyset\}$ . At time  $n$ , the channel input is

$$X_n = A(\xi_{n-1})X_{n-1} + \frac{1}{g}B(\xi_{n-1})Z_{n-1}.$$

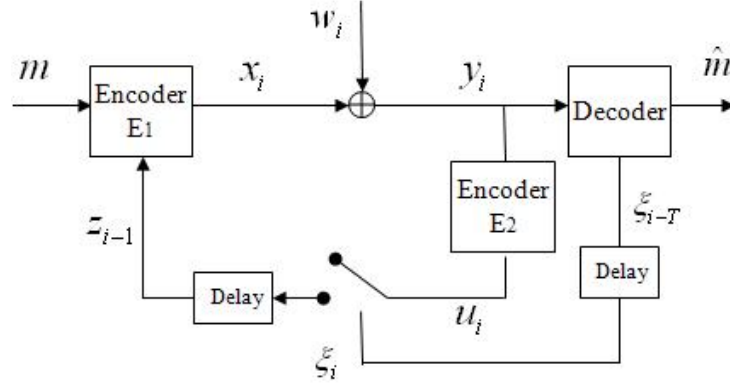


Figure 7.5 1st-MV Gaussian channel with intermittent feedback.

Now, we construct the decoder as follows. For the first  $T$  times, the decoder receives  $Y_1^T$  and the state of the decoder are set to be zero (i.e.  $\tilde{X}_0 = \tilde{X}_1 = \dots = \tilde{X}_{T-1} = 0$ ). At time  $n$  ( $n > T$ ), the decoder receives  $(Y_0^n, \xi_0^{n-T})$  and

$$\begin{aligned}\tilde{X}_n &= A(\xi_{n-T})\tilde{X}_{n-1} + B(\xi_{n-T})Y_{n-T}, \\ \hat{X}_{0,n} &= \left( \prod_{j=0}^{n-T} a(\xi_j) \right)^{-1} \tilde{X}_n.\end{aligned}$$

Now, we present a theorem below and give the proof in Appendix.

**Theorem 90** *Consider a first-order moving average Gaussian channel with random intermittent feedback where the intermittent acknowledgement is available to the decoder with finite delay (Fig. 7.5). Then the coding strategy  $\mathcal{C}$  has a transmission rate arbitrarily close to  $R = (1 - e) \log(|r|)$  with error probability decaying at least exponentially.*

**Remark 91** *As it is shown in the Theorem, the finite delay  $T$  does not affect the transmission rate. However, it will be shown in the proof that the delay actually affects the probability of decoding error.*

**Corollary 92** *Assume the erasure probability  $e = 0$  (i.e. perfect feedback). Then for any  $\epsilon > 0$ , there exists a  $\alpha^* > 0$  such that, for any 1-MV Gaussian channels with memory  $|\alpha| \leq \alpha^*$ , the*



coding strategy  $\mathcal{C}$  has a transmission rate  $R > C_{fb} - \epsilon$  where  $C_{fb}$  denotes the perfect feedback capacity.

**Remark 93** *The corollary implies that our linear coding scheme converges to the capacity-achieving coding scheme for the 1st-MV Gaussian channel with small  $\alpha$ . This will be verified later by a simulation result.*

Now we show the proof of Theorem 90.

**Proof.** (*sketch*) Recall the perfect feedback capacity for the 1st-MV Gaussian channel in Kim (2010) is  $\log(r)$  where  $r$  is the unique positive root of the fourth-order polynomial

$$P_1 x^2 = (x^2 - 1)(x^2 + 2\lambda\alpha x + \alpha^2), \quad (7.8)$$

where

$$\lambda = \begin{cases} 1 & \alpha \leq 0 \\ -1 & \alpha > 0 \end{cases}.$$

After some algebra, we have the polynomial (7.7) be equivalent to

$$P_1 x^2 = (x^2 - 1)(x^2 + \alpha^2 x^2 + 2\alpha x). \quad (7.9)$$

Clearly, the polynomial (7.9) converges to (7.8) as  $\alpha$  goes to zero. Thus, their corresponding roots are close as well. The proof is complete. ■

**Corollary 94** *If the erasure probability  $e = 0$  (i.e. perfect feedback) and  $\alpha = 0$  (i.e. the additive Gaussian noise is white), the rate  $R$  achieves the capacity  $C = \frac{1}{2} \log(1 + P)$  and the coding scheme  $\mathcal{C}$  converges to the well-known Schalkwijk-Kailath scheme.*

**Proof.** (*sketch*) For  $\alpha = 0$ , the polynomial (7.7) becomes

$$x^3 - (1 + P)x = 0$$

and thus  $r = \sqrt{1 + P}$ . That is, the achievable rate of our proposed coding scheme  $R = \frac{1}{2} \log(1 + P)$  which is the capacity of the channel. It is straightforward to demonstrate that for  $\alpha = 0$  the coding scheme  $\mathcal{C}$  turns to be the Schalkwijk-Kailath scheme. ■

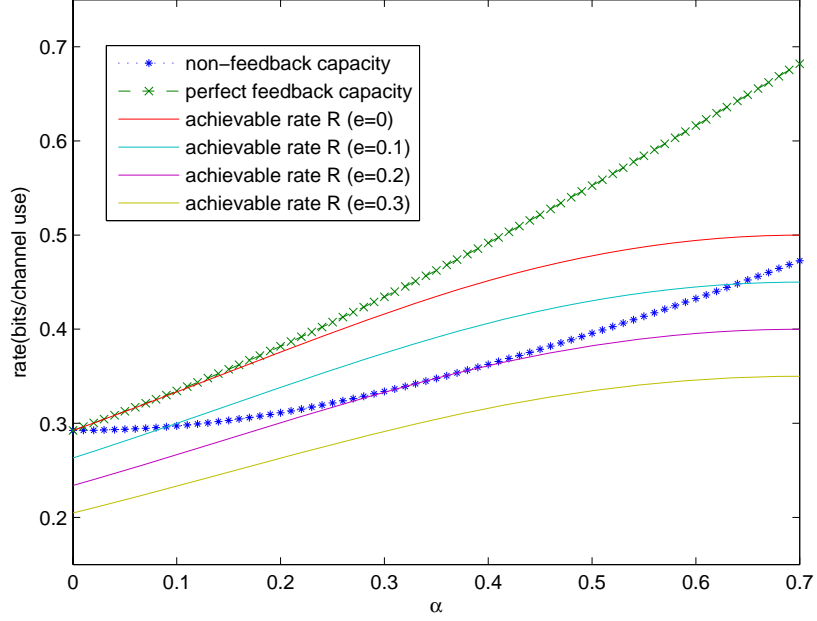


Figure 7.6 Achievable rates of coding scheme  $\mathcal{C}$  under power constraints  $P_1 = P_2 = 0.5$ .

We end this section by presenting a simulation result. See Fig. 7.6. It is shown that the achievable rate of the proposed linear coding scheme is almost equal to the perfect feedback capacity when the erasure probability  $e = 0$  and  $\alpha \in [0, 0.15]$ . This verifies Corollary 92. As  $e$  increases, the achievable rate decreases proportionally. For  $e < 0.2$ , there exists a range of  $\alpha$  such that the achievable rate is larger than the non-feedback capacity. For  $e > 0.2$ , although the transmission rate is still positive, the rate enhancement (with respect to the non-feedback capacity) disappears.

## 7.5 Conclusion

In this chapter, we have introduced and investigated a new framework of noisy feedback communication channels with side information at the decoder. Firstly, we have shown that its capacity can be characterized by the causal conditioning directed information. As an example, the n-block capacity of Gaussian noisy feedback channels under this framework has been char-

acterized and an iteration algorithm has been proposed to obtain a lower bound. Secondly, we have considered a class of noisy feedback channels — 1st-MV Gaussian channels with intermittent feedback. We have proposed a linear coding strategy which, for certain cases, provides a transmission rate larger than the non-feedback capacity. This implies that the our new framework allows simple linear coding strategies with significant positive transmission rate.

The new framework opens many promising avenues for future research. We briefly list a few of them as follows.

1. The capacity of noisy feedback channels with side information at the decoder is clearly an non-trivial upper bound on the capacity of classical noisy feedback channels. Thus, besides the Gaussian channel investigated in this chapter, we may be able to characterize the capacity for different classes of noisy feedback channels under our new framework in order to obtain tight upper bounds on the capacity of classical noisy feedback channels. For example, finite state channel is a good candidate to work with.
2. It is of importance to find calculation approaches for the causal conditioning directed information. We speculate that the calculation should be amenable to the approaches developed for the calculation of the directed information, e.g. dynamic programming (Tatikonda and Mitter (2009)). This, however, remains to be seen.
3. The new framework could be extended to the multi-terminal case and a fruitful results could be obtained. For example, by considering our new framework, it is hopeful to find an outer bound for multiple access channels with noisy feedback.

## CHAPTER 8. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

### 8.1 Conclusion

Motivated by the emerging networked system with interconnecting communications, a comprehensive mathematical theory of communication channels with noisy feedback is much more stringent than any time before. However, a careful review on the literature finds out that very few results have been derived on this subject, due to the intractable coordination loss between the transmitter and the receiver.

This thesis serves as a step forward to complete this mathematical theory. First of all, we have analyzed the information flow in communication channels with noisy feedback. An *information flow decomposition equality* has been derived and served as a basis for the rest of the results in this thesis. In addition, we have proposed a new concept, *residual directed information*, which is equal to the mutual information between the message and the channel outputs. This new concept indicates that the well-known mutual information and the directed information are not suitable for characterizing the message-delivery information flow in channels with access to noisy feedback. With the the information flow equality and the new concept in hand, we have developed multiple novel results on noisy feedback systems:

1. We provided a channel coding theorem, characterized by the residual directed information, for finite-alphabet communication channels with noisy feedback. Then we derived upper and lower bounds on the capacity, which are characterized by the causal conditioning directed information.
2. We derived upper and lower bounds on the n-block capacity of Gaussian channels with additive Gaussian noise feedback. These bounds can be numerically obtained by solving

well-defined convex optimization problems. Under the assumption of stationarity on Gaussian noises, we proved that the limits of these bounds exist and can be characterized in the form of power spectral optimizations.

3. We provided a necessary condition on the capacity-achieving channel codes for *feedback-unfavorable* channels (i.e. a class of channels whose capacity can not be increase by using feedback). Based on this condition, we concluded that any capacity-achieving channel code for *feedback-unfavorable* channels could not use feedback information.
4. Finally, we investigated an extended noisy feedback setting - noisy feedback communications with side information at the decoder, where the feedback information received by the transmitter is also available to the decoder with some finite delays. We proved that the capacity of this class of channels can be characterized by the causal conditional directed information; for additive Gaussian noises, the new framework allows linear feedback coding schemes with positive transmission rate, which is (in certain regime) much larger than the non-feedback Gaussian channel capacity.

In summary, we have proved that the information flow decomposition equality is a foundation and powerful tool to deal with noisy feedback problems. We anticipate that this equality will serve as a basis for more valuable results and be helpful in establishing the comprehensive theory of communications with noisy feedback.

## 8.2 Future Directions

We list several avenues for future research, which are suggested by the main results in this thesis.

### 8.2.1 Noisy feedback capacity and computable bounds

As it is shown, the capacity characterization in Theorem 31 is not computable in general due to the probabilistic limit and code-functions. We simplified this characterization under certain reasonable conditions (e.g. strong converse property). However, is it possible to obtain a single-letter characterization for some particular channels with noisy feedback?

From industrial point of view, some easy-computable bounds on the noisy feedback capacity are vital reference for communication product analysis and design. Therefore, besides the Gaussian channel considered in this thesis, we wish to find out computable bounds on the capacity for other important/common channels.

### 8.2.2 To use or not to use feedback

Theorem 73 implies that there does not exist a capacity-achieving feedback coding scheme for DMC with noisy feedback. In other words, exploiting the information from the noisy feedback link is actually detrimental to achieving the maximal achievable rate. This negative result induces a natural question: to use or not to use feedback. To answer this question, there are three research directions that become relevant in this context, and that propose to investigate.

1. Theorem 73 shows that using noisy feedback inevitably causes rate-loss. However, as proved in the literature, using noisy feedback is still beneficial to improve decoding error exponent (Draper and Sahai (2006a,b, 2008)), simplify coding structure (Agrawal and Love (2011)), etc. Therefore, we need a tradeoff while using noisy feedback. How to characterize the tradeoff, however, is an important issue. In particular, what is the mathematical relationship between the rate-loss and the improved error exponent/coding complexity? We believe that this tradeoff characterization will perform as a basic reference for communication engineer to do physical layer design.
2. Finally, as we move away from pure communication problems and consider problems like remote feedback stabilization, we have no choice but to use feedback (from noisy DMC channels). In these situations, the loss of communication efficiency needs to be contrasted and compared with the benefits of using feedback. It is necessary to derive this formal analysis. The starting point will be the equivalence between stabilization and communication with noiseless feedback for Gaussian channels (Elia (2004)). When the feedback is noisy, this equivalence seems to partially break down. While there are still enough bits reliably flowing back to the transmitter to guarantee stabilization, it is no longer immediate to identify a suitable decoder at the receiver side. The information

pattern in this case is different and maybe the existence of such decoder is questionable.

### 8.2.3 Explicit linear feedback coding schemes

One of the long-term objective is to find out explicit capacity-achieving coding schemes for noisy feedback communication systems. Motivated by the CP-scheme, we will begin with designing the best linear feedback coding scheme. Note, however, that for noisy feedback communication Kim et al. (2007) has proved that any coding scheme linearly using both the message index and the feedback information can not achieve any positive transmission rate. Therefore, the “linear coding” herein refers to the linear processing of the feedback information only. Although it may not achieve the capacity, the linear scheme will play an important role in academic field and certainly have widely industrial applications.

The first effort will be devoted into additive colored Gaussian channels with additive Gaussian noise feedback. One possible approach we will investigate is to extend the Cover-Pombra (CP) scheme which is a capacity-achieving scheme for noiseless feedback Gaussian channels.

As it is known, CP scheme allows Gaussian signalling of the message information and linear processing of the feedback information<sup>1</sup>. Furthermore, CP scheme implicitly incorporates the Kalman filtering algorithm and therefore is a “hub” connecting communications and estimation. To extend the CP scheme or its code construction idea, we need answer several necessary questions summarized below.

1. Is it possible to construct a CP-like feedback coding scheme for noisy feedback Gaussian channels?
2. What is the rate-loss (the gap between the capacity and the achievable rate of CP-like scheme) and the decaying rate of the probability of decoding error?
3. As did in the noiseless feedback case, does the constructed CP-like scheme lead to a convergence of communications and estimation?

If possible, it is worthy to investigate the best linear coding scheme for other important channels and in the meanwhile try to find the capacity-achieving feedback coding scheme.

---

<sup>1</sup>This configuration can be alternatively viewed as a concatenated coding scheme.

### 8.2.4 Multi-terminal communications with noisy feedback

We anticipate that the information flow decomposition equality is extendable to multi-terminal settings and helpful to find outer/inner bounds on the capacity region of multi-access channel (MAC), broadcast channel (BC), relay channel (RC) under the situation of noisy feedback. This work will serve as a basis for further investigation of general network capacity with node-to-node noisy feedback communication. So far, quite few researchers have been working on this problem due to the lack of theoretic results for point-to-point noisy feedback communication systems. See Gastpar and Kramer (2006); Lapidoth and Wigger (2010) and reference therein.

1. One conventional approach to obtain achievable region (inner bound) is to design specific coding schemes. The effective (linear) coding structures obtained under the point-to-point noisy feedback communication will be strong candidates to investigate. In particular, the CP-like coding structure or the concatenated coding structure may be applicable to the multi-terminal case. Moreover, some conventional coding techniques will be tested, such as superposition coding, simultaneous non-unique decoding, etc.
2. The outer bound is mostly obtained by invoking Fano's inequality, dependence-balance argument, etc. Besides, extend the computable upper bound derived for the point-to-point noisy feedback communication to the multi-terminal case will be a valuable attempt.

### 8.2.5 Confluence of feedback control and feedback communication

One important research avenue is to further develop an integrated view and theory of feedback control and feedback communication. For noiseless feedback and Gaussian channels, previous work Elia (2004); Liu and Elia (2006) has shown the equivalence between communication with feedback and stabilization with feedback, and the convergence of fundamental limitations of communication control and estimation for feedback systems. In particular, it has shown that a noiseless feedback system must satisfy the Bode Integral formula, a fundamental limitation of feedback systems. at the same time, the Bode integral must be equal to the communication rate in the channel, which is expressed by the directed information. Finally, both quantities are equal to the rate of the decay of the Krame-Rao bound of an equivalent



estimation problem, showing the necessity of a Kalman filter in the decoder. The relations are summarized as follows:

$$Rate = I(W; \hat{W}) = I(U \rightarrow Y) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log |\mathcal{S}(e^{j2\pi\theta})| d\theta = - \lim_{T \rightarrow \infty} \frac{CRB_{W,T}}{2(T+1)} = \sum_i \log \lambda_i^u(A)$$

where  $\mathcal{S}(e^{j2\pi\theta})$  is the Sensitivity transfer function from  $N$  to  $y$ ,  $CRB_{W,t} \leq \mathbf{E}\{(W - \hat{W}_t)(W - \hat{W}_t)'\}$  is the Cramer-Rao bound, and  $\lambda_i^u(A)$  are the unstable eigenvalues of  $A$ .

This unified theory allows us to use results, controller design tools and methodologies to deal with feedback communication problems (e.g. coding design, system analysis, etc) and *vice versa*. For the noiseless feedback case, much work has been done and many notable results have been obtained. In particular, we can use information theoretic quantities, i.e. directed information to characterize limitation of feedback systems over communication channels. This idea has been further explored in Martins and Dahleh (2008), where certain feedback performance measure of disturbance rejection was expressed in terms of the directed information. Unfortunately, the above results heavily depend on noiseless feedback. For the noisy feedback case, however, only few results (Yuksel and Bassar (2011)) have been known and much work remains to be done. Now that we have an appropriate information theoretic quantity characterizing the noisy feedback rate and capacity, it is then necessary to re-investigate the equivalence and connections with control and estimation settings. we remark that the intuition to the insight that lead to the introduction of the residual directed information was obtained from the control and communication analysis of a typical feedback control system with disturbance and sensor noise (Elia (2005)). This encourages us to solve the noisy feedback communication problem from feedback control perspective. As an illustration, it is known that the most general linear controller has two degree of freedom, one is in the feedback and the other is feed-forward. We plan to explore the two-degree controller synthesis methodology to the coding design for communications systems. However, it is reasonable to believe that the two-degree coding design would in general lead to a non-convex problem. At this point, the duality theory may provide a solvable result or at least a sub-optimal solution.

## APPENDIX A. PROOFS OF RESULTS IN CHAPTER 3

### Proof of Proposition 20

1) By Corollary 19, we have

$$\begin{aligned}
 \mathbb{E}[i^R(X^n(M) \rightarrow Y^n)] &= I^R(X^n(M) \rightarrow Y^n) \\
 &\leq I(X^n; Y^n) \\
 &= H(Y^n) - H(Y^n|X^n) \\
 &\leq H(Y^n) \\
 &\leq \log |\mathcal{Y}^n|
 \end{aligned}$$

2)

$$\begin{aligned}
 i^R(X^n(M) \rightarrow Y^n) &= i(X^n \rightarrow Y^n) - i(X^n \rightarrow Y^n|M) \\
 &= \log \frac{\vec{p}(Y^n|X^n)}{p(Y^n)} - \log \frac{\vec{p}(Y^n|X^n, M)}{p(Y^n|M)} \\
 &= \log \frac{\vec{p}(Y^n|X^n)}{p(Y^n)} - \log \frac{\vec{p}(Y^n|X^n)}{p(Y^n|M)} \quad (a) \\
 &= \log \frac{p(Y^n|M)}{p(Y^n)}
 \end{aligned}$$

where line (a) follows the fact that  $\mathcal{M} \rightarrow X^n \rightarrow Y^n$  forms a Markov chain. Then,

$$\begin{aligned}
& \text{Var}[i^R(X^n(M) \rightarrow Y^n)] \\
&= \mathbb{E}[(i^R(X^n(M) \rightarrow Y^n))^2] - \mathbb{E}^2[i^R(X^n(M) \rightarrow Y^n)] \\
&\leq \mathbb{E}[(i^R(X^n(M) \rightarrow Y^n))^2] \\
&= \mathbb{E}[(\log \frac{p(Y^n|M)}{p(Y^n)})^2] \\
&= \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n} p(m, y^n) (\log \frac{p(y^n|m)}{p(y^n)})^2 \\
&= \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n, p(m, y^n) \geq p(y^n)} p(m, y^n) (\log \frac{p(y^n|m)}{p(y^n)})^2 \\
&\quad + \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n, p(m, y^n) \leq p(y^n)} p(m, y^n) (\log \frac{p(y^n|m)}{p(y^n)})^2 \\
&= \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n, p(m, y^n) \geq p(y^n)} p(m, y^n) (\log \frac{p(y^n)}{p(y^n|m)})^2 \\
&\quad + \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n, p(m, y^n) \leq p(y^n)} p(m, y^n) (\log \frac{p(y^n|m)}{p(y^n)})^2 \\
&\leq \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n, p(m, y^n) \geq p(y^n)} p(m, y^n) (\log \frac{1}{p(y^n|m)})^2 \\
&\quad + \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n, p(m, y^n) \leq p(y^n)} p(m, y^n) (\log \frac{1}{p(y^n)})^2 \\
&\leq \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n} p(m, y^n) (\log \frac{1}{p(y^n|m)})^2 + \sum_{m \in \mathcal{M}, y^n \in \mathcal{Y}^n} p(m, y^n) (\log \frac{1}{p(y^n)})^2 \\
&= \sum_{m \in \mathcal{M}} p(m) \sum_{y^n \in \mathcal{Y}^n} p(y^n|m) (\log \frac{1}{p(y^n|m)})^2 + \sum_{y^n \in \mathcal{Y}^n} p(y^n) (\log \frac{1}{p(y^n)})^2 \\
&\leq \sum_{m \in \mathcal{M}} p(m) \sum_{y^n \in \mathcal{Y}^n} 1 + \sum_{y^n \in \mathcal{Y}^n} 1 \quad (b) \\
&= 2|\mathcal{Y}^n|
\end{aligned}$$

(b) follows the fact that function  $f(x) = x(\log(\frac{1}{x}))^2 \leq 1$  for  $0 \leq x \leq 1$ . This is easy to check by taking derivative on  $x$ .

### Proof of Corollary 22

We herein adopt a derivation methodology similar to the one used in Theorem 17.

$$\begin{aligned}
& I(M; Y^n) \\
&= H(Y^n) - H(Y^n | M) \\
&= \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M) \\
&= \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M, V^{i-1}) - \left( \sum_{i=1}^n H(Y_i | Y^{i-1}, M) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M, V^{i-1}) \right) \\
&\stackrel{(a)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M, Z^{i-1}) - \left( \sum_{i=1}^n H(Y_i | Y^{i-1}, M) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M, V^{i-1}) \right) \\
&= \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Y_i | Y^{i-1}, X^i) - \left( \sum_{i=1}^n H(Y_i | Y^{i-1}, M) - \sum_{i=1}^n H(Y_i | Y^{i-1}, M, V^{i-1}) \right) \\
&= \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}) - \sum_{i=1}^n I(V^{i-1}; Y_i | Y^{i-1}, M) \\
&= I(X^n \rightarrow Y^n) - I(V^{n-1} \rightarrow Y^n | M)
\end{aligned}$$

where line (a) follows from the fact that  $Z^{i-1} = Y^{i-1} + V^{i-1}$ . Next,

$$\begin{aligned}
I(V^{n-1} \rightarrow Y^n | M) &\stackrel{(b)}{=} I(V^{n-1}; Y^n | M) \\
&= H(V^{n-1} | W) - H(V^{n-1} | Y^n, M) \\
&\stackrel{(c)}{=} H(V^{n-1}) - H(V^{n-1} | Y^n) + H(V^{n-1} | Y^n) - H(V^{n-1} | Y^n, M) \\
&= I(V^{n-1}; Y^n) + I(M; V^{n-1} | Y^n)
\end{aligned}$$

where line (b) follows from the fact that there exists no feedback from  $Y^n$  to  $V^{n-1}$  and line (c) follows from the fact that the noise  $V^{n-1}$  is independent from  $M$ . Putting previous equations together, the proof is complete.

## APPENDIX B. PROOFS OF RESULTS IN CHAPTER 4

### Proof of Lemma 30

Before giving the proof, we need the following Lemma.

**Lemma 95** *For channels with noisy feedback, as shown in Fig.4.1,*

$$p(x^n, y^n) = \sum_{z^n \in \mathcal{Z}^n} \prod_{i=1}^n \underbrace{p(z_i | y^i, z^{i-1})}_{\text{Feedback link}} \underbrace{p(x_i | x^{i-1}, z^{i-1})}_{\text{Encoding}} \underbrace{p(y_i | x^i, y^{i-1})}_{\text{Channel}}$$

**Proof.**

$$\begin{aligned} p(x^n, y^n) &= \sum_{z^n \in \mathcal{Z}^n} p(x^n, y^n, z^n) \\ &= \sum_{z^n \in \mathcal{Z}^n} p(z_n | x^n, y^n, z^{n-1}) p(x^n, y^n, z^{n-1}) \\ &= \sum_{z^n \in \mathcal{Z}^n} p(z_n | x^n, y^n, z^{n-1}) p(y_n | x^n, y^{n-1}, z^{n-1}) p(x^n, y^{n-1}, z^{n-1}) \\ &= \sum_{z^n \in \mathcal{Z}^n} p(z_n | x^n, y^n, z^{n-1}) p(y_n | x^n, y^{n-1}, z^{n-1}) p(x_n | x^{n-1}, y^{n-1}, z^{n-1}) \\ &\quad p(x^{n-1}, y^{n-1}, z^{n-1}) \\ &\stackrel{(a)}{=} \sum_{z^n \in \mathcal{Z}^n} p(z_n | y^n, z^{n-1}) p(y_n | x^n, y^{n-1}) p(x_n | x^{n-1}, z^{n-1}) \\ &\quad p(x^{n-1}, y^{n-1}, z^{n-1}) \\ &= \sum_{z^n \in \mathcal{Z}^n} \prod_{i=1}^n p(z_i | y^i, z^{i-1}) p(x_i | x^{i-1}, z^{i-1}) p(y_i | x^i, y^{i-1}) \end{aligned}$$

where (a) follows from the Markov chains:  $x^n - (y^n, z^{n-1}) - z_n$ ,  $z^{n-1} - (x^n, y^{n-1}) - y_n$  and  $y^{n-1} - (x^{n-1}, z^{n-1}) - x_n$ . ■

Now, we are ready to give the proof of Lemma 30. **Proof.**

$$\begin{aligned}
& p(x^n, y^n, f^n) \\
&= p(x^n, y^n | f^n) p(f^n) \\
&\stackrel{(a)}{=} p(f^n) \sum_{z^n \in \mathcal{Z}^n} \prod_{i=1}^n p(z_i | y^i, z^{i-1}, f^n) p(x_i | x^{i-1}, z^{i-1}, f^n) p(y_i | x^i, y^{i-1}, f^n) \\
&= p(f^n) \sum_{z^n \in \{ \mathcal{Z}^n : x^n = f^n(z^{n-1}) \}} \prod_{i=1}^n p(z_i | y^i, z^{i-1}, f^n) p(y_i | f^i(z^{i-1}), y^{i-1}, f^n) \\
&\stackrel{(b)}{=} p(f^n) \sum_{z^n \in \{ \mathcal{Z}^n : x^n = f^n(z^{n-1}) \}} \prod_{i=1}^n p(z_i | y^i, z^{i-1}) p(y_i | f^i(z^{i-1}), y^{i-1}) \\
&\stackrel{(c)}{=} \prod_{i=1}^n \prod_{z^{i-1}} p(f_i(z^{i-1}) | f^{i-1}(z^{i-2}), z^{i-1}) \sum_{z^n \in \{ \mathcal{Z}^n : x^n = f^n(z^{n-1}) \}} \prod_{i=1}^n p(z_i | y^i, z^{i-1}) p(y_i | f^i(z^{i-1}), y^{i-1})
\end{aligned}$$

where (a) follows from Lemma 95. Line (b) follows from the Markov chains:  $f^n - (y^i, z^{i-1}) - z_i$  and  $f^n - (f^i(z^{i-1}), y^{i-1}) - y_i$ . Line (c) follows from Lemma 29. ■

**Proof of Lemma 35**

$$\begin{aligned}
& I^R(X^n(F^n) \rightarrow Y^n) \\
& \stackrel{(a)}{=} I(F^n; Y^n) \\
& = I(F^n; (Y^n, Z^{n-1}) - I(F^n; Z^{n-1}|Y^n) \\
& \stackrel{(b)}{=} I(F^n \rightarrow (Y^n, Z^{n-1})) - I(F^n; Z^{n-1}|Y^n) \\
& = \sum_{i=1}^n I(F^i, (Y_i, Z_i)|Y^{i-1}, Z^{i-1}) - I(F^n; Z^{n-1}|Y^n) \\
& = \sum_{i=1}^n H(Y_i, Z_i|Y^{i-1}, Z^{i-1}) - H(Y_i, Z_i|Y^{i-1}, Z^{i-1}, F^i) - I(F^n; Z^{n-1}|Y^n) \\
& = \sum_{i=1}^n H(Z_i|Y^i, Z^{i-1}) + H(Y_i|Y^{i-1}, Z^{i-1}) - H(Z_i|Y^i, Z^{i-1}, F^i) - H(Y_i|Y^{i-1}, Z^{i-1}, F^i) \\
& \quad - I(F^n; Z^{n-1}|Y^n) \\
& \stackrel{(c)}{=} \sum_{i=1}^n H(Y_i|Y^{i-1}, Z^{i-1}) - H(Y_i|Y^{i-1}, Z^{i-1}, F^i) - I(F^n; Z^{n-1}|Y^n) \\
& \stackrel{(d)}{=} \sum_{i=1}^n H(Y_i|Y^{i-1}, Z^{i-1}) - H(Y_i|Y^{i-1}, X^i, Z^{i-1}, F^i) - I(F^n; Z^{n-1}|Y^n) \\
& \stackrel{(e)}{=} \sum_{i=1}^n H(Y_i|Y^{i-1}, Z^{i-1}) - H(Y_i|Y^{i-1}, X^i, Z^{i-1}) - I(F^n; Z^{n-1}|Y^n) \\
& = \sum_{i=1}^n I(X^i, Y_i|Y^{i-1}, Z^{i-1}) - I(F^n; Z^{n-1}|Y^n) \\
& = I(X^n \rightarrow Y^n || Z^{n-1}) - I(F^n; Z^{n-1}|Y^n)
\end{aligned}$$

where (a) follows from Lemma 28. Line (b) follows from the fact that there exists no feedback from  $(Y^n, Z^n)$  to  $F^n$  and thus the mutual information and directed information coincide. Line (c) follows from the fact that  $H(Z_i|Y^i, Z^{i-1}) = H(Z_i|Y^i, Z^{i-1}, F^i)$  since  $F^i - (Y^i, Z^{i-1}) - Z_i$  forms a Markov chain. Line (d) follows from the fact that  $X^i$  can be determined by  $F^i$  and the outputs of the feedback link  $Z^{i-1}$ . Line (e) follows from the Markov chain  $F^i - (Y^{i-1}, X^i, Z^{i-1}) - Y_i$ .

**Proof of Lemma 40**

$$\begin{aligned}
& |I(X^n \rightarrow Y^n | Z^{n-1}) - I(X^n \rightarrow Y^n || Z^{n-1} | S)| \\
&= \left| \sum_{i=1}^n I(X^i; Y_i | Z^{i-1}) - \sum_{i=1}^n I(X^i; Y_i | Z^{i-1}, S) \right| \\
&= \left| \sum_{i=1}^n H(Y_i | Y^{i-1}, Z^{i-1}) - H(Y_i | X^i, Y^{i-1}, Z^{i-1}) \right. \\
&\quad \left. - H(Y_i | Y^{i-1}, Z^{i-1}, S) - H(Y_i | X^i, Y^{i-1}, Z^{i-1}, S) \right| \\
&= \left| \sum_{i=1}^n I(S; Y_i | Y^{i-1}, Z^{i-1}) - \sum_{i=1}^n I(S; Y_i | X^i, Y^{i-1}, Z^{i-1}) \right| \\
&\leq \max \left\{ \sum_{i=1}^n I(S; Y_i | Y^{i-1}, Z^{i-1}), \sum_{i=1}^n I(S; Y_i | X^i, Y^{i-1}, Z^{i-1}) \right\} \\
&\leq \max \left\{ \sum_{i=1}^n I(S; Y_i, Z_i | Y^{i-1}, Z^{i-1}), \sum_{i=1}^n I(S; Y_i, X_{i+1}, Z_i | X^i, Y^{i-1}, Z^{i-1}) \right\} \\
&= \max \{ I(S; Y^n, Z^n), I(S; Y^n, X_2^{n+1}, Z^n) \} \\
&\leq \max \{ H(S), H(S) \} \\
&\leq \log |\mathcal{S}|
\end{aligned}$$

**Proof of Lemma 41**

$$\begin{aligned}
& p(x^N, y^N, z^{N-1}, s^n) \\
&= \sum_{s_{n+1}, \dots, s_N} p(x^N, y^N, z^{N-1}, s^N) \\
&= \sum_{s_{n+1}, \dots, s_N} \prod_{j=n+1}^N p(x_j, y_j, z_{j-1}, s_j | x^{j-1}, y^{j-1}, z^{j-2}, s^{j-1}) \cdot p(x^n, y^n, z^{n-1}, s^n) \\
&= \sum_{s_{n+1}, \dots, s_N} \prod_{j=n+1}^N p(x_j | x^{j-1}, y^{j-1}, z^{j-1}, s^{j-1}) \cdot p(y_j, s_j | x^j, y^{j-1}, z^{j-1}, s^{j-1}) \\
&\quad \cdot p(z_{j-1} | x^{j-1}, y^{j-1}, z^{j-2}, s^{j-1}) \cdot p(x^n, y^n, z^{n-1}, s^n) \\
&= \sum_{s_{n+1}, \dots, s_N} \prod_{j=n+1}^N p(x_j | x^{j-1}, z^{j-1}) \cdot p(y_j, s_j | x_j, s_{j-1}) \cdot p(z_{j-1} | y^{j-1}, z^{j-2}) \cdot p(x^n, y^n, z^{n-1}, s^n)
\end{aligned}$$



where the last line follows from the channel causality models. Similarly,

$$\begin{aligned}
& p(x^N, y^{N-1}, z^{N-1}, s^n) \\
&= \sum_{s_{n+1}, \dots, s_N} \sum_{y_N} p(x^N, y^N, z^{N-1}, s^N) \\
&= \sum_{s_{n+1}, \dots, s_N} \sum_{y_N} \prod_{j=n+1}^N p(x_j | x^{j-1}, z^{j-1}) \cdot p(y_j, s_j | x_j, s_{j-1}) \\
&\quad \cdot p(z_{j-1} | y^{j-1}, z^{j-2}) \cdot p(x^n, y^n, z^{n-1}, s^n) \\
&= \sum_{s_{n+1}, \dots, s_N} \prod_{j=n+1}^N p(x_j | x^{j-1}, z^{j-1}) \cdot p(z_{j-1} | y^{j-1}, z^{j-2}) \\
&\quad \prod_{j=n+1}^{N-1} p(y_j, s_j | x_j, s_{j-1}) \cdot p(s_N | x_N, s_{N-1}) \cdot p(x^n, y^n, z^{n-1}, s^n)
\end{aligned}$$

Next,

$$\begin{aligned}
p(y_N | x^N, y^{N-1}, z^{N-1}, s^n) &= \frac{p(x^N, y^N, z^{N-1}, s^n)}{p(x^N, y^{N-1}, z^{N-1}, s^n)} \\
&= \frac{\sum_{s_{n+1}, \dots, s_N} \prod_{j=n+1}^N p(y_j, s_j | x_j, s_{j-1})}{\sum_{s_{n+1}, \dots, s_N} \prod_{j=n+1}^{N-1} p(y_j, s_j | x_j, s_{j-1}) \cdot p(s_N | x_N, s_{N-1})}
\end{aligned}$$

It is observed that  $(x^n, y^n, s^{n-1}, z^{n-1})$  does not appear in the last line, we conclude that

$$p(y_N | x^N, y^{N-1}, z^{N-1}, s^n) = p(y_N | x_{n+1}^N, y_{n+1}^{N-1}, z_n^{N-1}, s_n)$$

## APPENDIX C. PROOFS OF RESULTS IN CHAPTER 5

### Proof of Lemma 51

#### *i).proof of equality (1)*

Since there is no feedback from channel outputs  $Y^n$  to the feedback additive noise  $V^{n-1}$ , the directed information from  $V^{n-1}$  to  $Y^n$  equals the mutual information. Thus, we have

$$\begin{aligned}
 I(V^{n-1}; Y^n) &= I(V^{n-1} \rightarrow Y^n) \\
 &= \sum_{i=1}^n I(V^{i-1}; Y_i | Y^{i-1}) \\
 &= \sum_{i=1}^n h(Y_i | Y^{i-1}) - h(Y_i | Y^{i-1}, V^{i-1}) \\
 &= h(Y^n) - h(Y^n | V^{n-1})
 \end{aligned}$$

In addition,

$$I(V^{n-1}; Y^n) = h(Y^n) - h(Y^n | V^{n-1}).$$

Thus, we have  $h(Y^n | V^{n-1}) = h(Y^n | V^{n-1})$ .

#### *ii).proof of equality (2)*

Recall the chain rule of conditional mutual information as follows,

$$I(a, b; c | d) = I(b; c | d) + I(a; c | b, d), \tag{C.1}$$

where  $(a, b, c, d)$  refers to random variables or vectors. First of all, we have

$$\begin{aligned}
& I(X^n \rightarrow Y^n || V^{n-1}) \\
&= \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}, V^{i-1}) \\
&= \sum_{i=1}^n I(M, X^i; Y_i | Y^{i-1}, V^{i-1}) - I(M; Y_i | X^i, Y^{i-1}, V^{i-1}) \\
&\stackrel{(a)}{=} \sum_{i=1}^n I(M, X^i; Y_i | Y^{i-1}, V^{i-1}) \\
&\stackrel{(b)}{=} \sum_{i=1}^n I(M; Y_i | Y^{i-1}, V^{i-1}) \\
&\stackrel{(c)}{=} \sum_{i=1}^n I(M, V_i^{n-1}; Y_i | Y^{i-1}, V^{i-1}) - I(V_i^{n-1}; Y_i | M, Y^{i-1}, V^{i-1})
\end{aligned}$$

where line (a) follows from the channel causality, i.e.,  $p(y_i | x^i, y^{i-1}, v^{i-1}, m) = p(y_i | x^i, y^{i-1})$ , and thus  $I(M; Y_i | X^i, Y^{i-1}, V^{i-1}) = 0$ , line (b) follows from the fact that  $X^i$  is a deterministic function of  $(M, Y^{i-1} + V^{i-1})$  and line (c) follows from the chain rule (C.1).

Now we show  $I(V_i^{n-1}; Y_i | M, Y^{i-1}, V^{i-1}) = 0$ . By using chain rule (C.1), we have

$$\begin{aligned}
I(V_i^{n-1}; Y_i | M, Y^{i-1}, V^{i-1}) &= I(V_i^{n-1}; Y^i, M | V^{i-1}) \\
&\quad - I(V_i^{n-1}; Y^{i-1}, M | V^{i-1})
\end{aligned}$$

Now, we have two facts: *i)* Given  $V^{i-1}$ ,  $(Y^i, M)$  is a deterministic function of  $(W^i, M)$ ; *ii)* the Markov chain  $V_i^{n-1} - V^{i-1} - (M, W^i)$  holds. Thus, the Markov chain  $V_i^{n-1} - V^{i-1} - (Y^i, M)$  holds. Consequently,

$$I(V_i^{n-1}; Y^i, M | V^{i-1}) = I(V_i^{n-1}; Y^{i-1}, M | V^{i-1}) = 0,$$

and, thus,  $I(V_i^{n-1}; Y_i | M, Y^{i-1}, V^{i-1}) = 0$ .

Proceed the above derivation, we have

$$\begin{aligned}
& I(X^n \rightarrow Y^n || V^{n-1}) \\
&= \sum_{i=1}^n I(M, V_i^{n-1}; Y_i | Y^{i-1}, V^{i-1}) \\
&= \sum_{i=1}^n I(M; Y_i | Y^{i-1}, V^{n-1}) + I(V_i^{n-1}; Y_i | M, Y^{i-1}, V^{i-1})
\end{aligned}$$

The second term is zero due to the fact that

$$I(V_i^{n-1}; Y_i | Y^{i-1}, V^{i-1}) = I(V_i^{n-1}; Y_i | V^{i-1}) - I(V_i^{n-1}; Y^{i-1} | V^{i-1})$$

and

$$I(V_i^{n-1}; Y^{i-1} | V^{i-1}) \leq I(V_i^{n-1}; Y_i | V^{i-1}) \leq I(V_i^{n-1}; Y_i, M | V^{i-1}) = 0.$$

where the last equality follows from (C). Therefore, we obtain

$$\begin{aligned} & I(X^n \rightarrow Y^n | V^{n-1}) \\ &= \sum_{i=1}^n I(M, ; Y_i | Y^{i-1}, V^{n-1}) \\ &= \sum_{i=1}^n h(Y_i | Y^{i-1}, V^{n-1}) - h(Y_i | M, Y^{i-1}, V^{n-1}) \\ &= \sum_{i=1}^n h(Y_i | Y^{i-1}, V^{n-1}) - h(Y_i | M, Y^{i-1}, X^i(M, Y^{i-1} + V^{i-1}), V^{n-1}) \\ &\stackrel{(a)}{=} \sum_{i=1}^n h(Y_i | Y^{i-1}, V^{n-1}) - h(Y_i | Y^{i-1}, X^i, V^{n-1}) \\ &= I(X^n \rightarrow Y^n | V^{n-1}) \end{aligned}$$

where line (a) follows from the channel causality.

### Proof of Proposition 54

Let  $\mathbf{H}_n = (\mathbf{I}_n + \mathbf{B}_n)\mathbf{K}_{w,n}(\mathbf{I}_n + \mathbf{B}_n)^T + \mathbf{K}_{s,n} + \mathbf{B}_n\mathbf{K}_{v,n}\mathbf{B}_n^T$ , we have

$$\begin{aligned} & \frac{1}{2} \log \frac{\det((\mathbf{I}_n + \mathbf{B}_n)\mathbf{K}_{w,n}(\mathbf{I}_n + \mathbf{B}_n)^T + \mathbf{K}_{s,n})}{\det \mathbf{K}_{w,n}} \\ &= \frac{1}{2} \log \frac{\det(\mathbf{H}_n - \mathbf{B}_n\mathbf{K}_{v,n}\mathbf{B}_n^T)}{\det \mathbf{K}_{w,n}}. \end{aligned}$$

Next,

$$\begin{aligned} \text{tr}(\mathbf{K}_{x,n}) \leq nP &\Leftrightarrow \text{tr}(\mathbf{K}_{s,n} + \mathbf{B}_n(\mathbf{K}_{v,n} + \mathbf{K}_{w,n})\mathbf{B}_n^T) \leq nP \\ &\Leftrightarrow \text{tr}(\mathbf{H}_n - \mathbf{K}_{w,n}\mathbf{B}_n^T - \mathbf{B}_n\mathbf{K}_{w,n} - \mathbf{K}_{w,n}) \leq nP. \end{aligned}$$

By applying the Schur complement, we have the following equivalences.

1.

$$\det \begin{bmatrix} \mathbf{K}_{v,n}^{-1} & \mathbf{B}_n^T \\ \mathbf{B}_n & \mathbf{H}_n \end{bmatrix} = \det(\mathbf{H}_n - \mathbf{B}_n\mathbf{K}_{v,n}\mathbf{B}_n^T) \det \mathbf{K}_{v,n}^{-1}.$$

2.

$$\begin{aligned} \mathbf{K}_{s,n} \geq 0 &\Leftrightarrow \mathbf{H}_n - (\mathbf{I}_n + \mathbf{B}_n)\mathbf{K}_{w,n}(\mathbf{I}_n + \mathbf{B}_n)^T - \mathbf{B}_n\mathbf{K}_{v,n}\mathbf{B}_n^T \geq 0 \\ &\Leftrightarrow \begin{bmatrix} \mathbf{H}_n & \mathbf{I}_n + \mathbf{B}_n^T & \mathbf{B}_n^T \\ \mathbf{I}_n + \mathbf{B}_n & \mathbf{K}_{w,n}^{-1} & \mathbf{0}_n \\ \mathbf{B}_n & \mathbf{0}_n & \mathbf{K}_{v,n}^{-1} \end{bmatrix} \geq 0 \end{aligned}$$

By taking simple replacements on the original formula, the proof is complete.

### Proof of Proposition 55

According to ?), we apply interior-point algorithms to solve this LMI problem, which have a polynomial-time complexity. In particular, the number of operations to obtain a  $\delta$ -accurate solution is upper bounded by

$$MN^3 \log\left(\frac{V}{\delta}\right)$$

where  $M$  is the total row size of the LMI system,  $N$  is the total number of scalar decision variables, and  $V$  is a data-dependent scaling factor. In Corollary 54, we have  $M = 3n + 1$  and  $N = \frac{3}{2}n^2 - \frac{1}{2}n$ . The result is then straightforward to obtain.

### Proof of Theorem 58

As the proof follows the same approach in Kim (2010), we herein streamline the proof and concentrate on the main difference. Define  $\tilde{C}_{fb}^{noisy}$  as formula (5.13). By the Szegő-Kolmogorov-Krein theorem, we have

$$\tilde{C}_{fb}^{noisy} = \sup_{\{X_i\}\text{-stationary}} h(\mathcal{Y}|\mathcal{V}) - h(\mathcal{W})$$

where the supremum is taken over all stationary Gaussian process  $\{X_i\}_{i=-\infty}^{\infty}$  of the form  $X_i = S_i + \sum_k b_k(W_{i-k} + V_{i-k})$  where  $\{S_i\}_{i=-\infty}^{\infty}$  is stationary and independent of  $(\{W_i\}_{i=-\infty}^{\infty}, \{V_i\}_{i=-\infty}^{\infty})$  such that  $\mathbb{E}[X_i^2] \leq P$ .

We first show that

$$\bar{C}_{fb,n}^{noisy} \leq \tilde{C}_{fb}^{noisy} \tag{C.2}$$

for all  $n$ . Fix  $n$  and assume  $(\mathbf{K}_{s,n}^*, \mathbf{B}_n^*)$  achieves  $\bar{C}_{fb,n}^{noisy}$ . Consider a block-wise white process  $\{S_i\}_{i=kn+1}^{(k+1)n}$ ,  $-\infty \leq k < \infty$ , independent and identically distributed according to  $\mathbb{N}_n(0, \mathbf{K}_{s,n}^*)$ . By following the steps in the proof of Lemma 59, it is straightforward to have

$$\bar{C}_{fb,n}^{noisy} \leq \frac{1}{kn} (h(Y_1^{kn} | V_1^{kn-1}) - h(W_1^{kn}))$$

for all  $k$ . Next, we use the same technical skill from Kim (2010) to show the inequality (C.2). Define the time-shifted process  $\{X_i(t)\}_{i=-\infty}^{\infty}$  where  $X_i(t) = X_{i+t}$ . Similarly define  $\{Y_i(t)\}_{i=-\infty}^{\infty}$ ,  $\{W_i(t)\}_{i=-\infty}^{\infty}$  and  $\{V_i(t)\}_{i=-\infty}^{\infty}$ . Introduce a random variable  $T$ , uniformly distributed over  $\{1, 2, 3, \dots, n\}$  and independent of everything else. Then it is easy to check that  $\{X_i(T), Y_i(T), W_i(T), V_i(T)\}_{i=-\infty}^{\infty}$  is jointly stationary. Next, we define  $\{\tilde{X}_i, \tilde{Y}_i, \tilde{W}_i, \tilde{V}_i\}_{i=-\infty}^{\infty}$  as a jointly Gaussian process with the same mean and autocorrelation as the stationary process  $\{X_i(T), Y_i(T), W_i(T), V_i(T)\}_{i=-\infty}^{\infty}$ . Thus,

$$\begin{aligned} \bar{C}_{fb,n}^{noisy} &\leq \frac{1}{kn} (h(Y_1^{kn}(T) | V_1^{kn-1}(T), T) - h(W_1^{kn}(T) | T)) \\ &\stackrel{(a)}{=} \frac{1}{kn} (h(Y_1^{kn}(T) | V_1^{kn-1}(T), T) - h(W_1^{kn})) \\ &\leq \frac{1}{kn} (h(Y_1^{kn}(T) | V_1^{kn-1}(T)) - h(W_1^{kn})) \\ &= \frac{1}{kn} (h(\tilde{Y}_1^{kn} | V_1^{kn-1}) - h(W_1^{kn})) \end{aligned}$$

where (a) follows from the stationarity assumption on  $\{W_i\}_{i=1}^{\infty}$ . Taking  $k \rightarrow \infty$ , we obtain

$$\bar{C}_{fb,n}^{noisy} \leq h(\tilde{\mathcal{Y}} | \mathcal{V}) - h(\mathcal{W}) \leq \tilde{C}_{fb}^{noisy}$$

We now show the main idea of proving the other direction. Given  $\epsilon > 0$ , we let  $\{\tilde{X}_i\}_{i=-\infty}^{\infty}$  achieve  $\tilde{C}_{fb}^{noisy} - \epsilon$ . Define the corresponding channel outputs as  $\{\tilde{Y}_i\}_{i=-\infty}^{\infty}$ . Then,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy} &= \liminf_{n \rightarrow \infty} \max_{\{X_i\}_{i=1}^n} \frac{1}{n} (h(Y_1^n | V_1^{n-1}) - h(W_1^n)) \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} (h(\tilde{Y}_1^n | V_1^{n-1}) - h(W_1^n)) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} (h(\tilde{Y}_1^n | V_1^{n-1}) - h(W_1^n)) \\ &= h(\tilde{\mathcal{Y}} | \mathcal{V}) - h(\mathcal{W}) \\ &= \tilde{C}_{fb}^{noisy} - \epsilon \end{aligned}$$

Taking  $\epsilon \rightarrow 0$ , we obtain  $\liminf_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy} \geq \tilde{C}_{fb}^{noisy}$ . The technical discussion on power constraint is identical to that in Kim (2010), so it is omitted here. Combined with inequality (C.2), we know that the limit of  $C_{fb,n}^{noisy}$  exists and  $\lim_{n \rightarrow \infty} \bar{C}_{fb,n}^{noisy} = \tilde{C}_{fb}^{noisy}$ .

### Proof of Lemma 59

First of all, we need the following lemma.

**Lemma 96** *Consider the CP-like coding scheme as shown in Fig.5.2,*

$$I(S^n; Y^n | V^n) = I(X^n \rightarrow Y^n | V^{n-1}) = h(Y^n | V^{n-1}) - h(W^n)$$

**Proof.**

$$\begin{aligned} & I(S^n; Y^n | V^n) \\ &= h(Y^n | V^n) - h(Y^n | S^n, V^n) \\ &\stackrel{(a)}{=} h(Y^n | V^{n-1}) - h(Y^n | S^n, V^n) \\ &= h(Y^n | V^{n-1}) - h(S^n + (\mathbf{I}_n + \mathbf{B}_n)W^n + \mathbf{B}_n V^n | S^n, V^n) \\ &\stackrel{(b)}{=} h(Y^n | V^{n-1}) - h((\mathbf{I}_n + \mathbf{B}_n)W^n) \\ &= h(Y^n | V^{n-1}) - h(W^n) \end{aligned}$$

where line (a) follows from the fact that  $Y^n$  does not depend on the feedback noise  $V_n$  due to the single step feedback delay. Line (b) follows from the fact that  $(S^n, W^n, V^n)$  are mutually independent. From Theorem 49, we have

$$I(X^n \rightarrow Y^n | V^{n-1}) = h(Y^n | V^{n-1}) - h(W^n).$$

The proof is complete. ■

We next present the necessary super-additive lemma below, the proof of which can be found in Appendix 4A ?).

**Lemma 97** (*Super-additive Sequence*) *Let  $a_N$ ,  $N = 1, 2, \dots, \infty$  be a bounded sequence of numbers. Assume that, for all  $1 \leq n < N$ ,*

$$na_n + (N - n)a_{N-n} \leq Na_N,$$

then

$$\lim_{N \rightarrow \infty} a_N = \sup_N a_N.$$

Based on this result, we need to show that the upper bound  $\bar{C}_{fb,n}^{noisy}$  on the n-block capacity is super-additive, namely,

$$n\bar{C}_{fb,n}^{noisy} + m\bar{C}_{fb,m}^{noisy} \leq (n+m)\bar{C}_{fb,n+m}^{noisy} \quad (\text{C.3})$$

for all  $n, m \geq 1$ . Now we present the proof of Lemma 59 as follows. **Proof.** Fix  $n$  and assume  $(\mathbf{K}_{s,n}^*, \mathbf{B}_n^*)$  achieves  $\bar{C}_{fb,n}^{noisy}$ . Similarly, fix  $m$  and assume  $(\mathbf{K}_{s,m}^*, \mathbf{B}_m^*)$  achieves  $\bar{C}_{fb,m}^{noisy}$ . Consider a process  $\{S_i\}_{i=1}^{n+m}$  that is independent of  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$ , and block-wise white with  $\{S_i\}_{i=1}^n \sim \mathbb{N}_n(0, \mathbf{K}_{s,n}^*)$  and  $\{S_i\}_{i=n+1}^{n+m} \sim \mathbb{N}_m(0, \mathbf{K}_{s,m}^*)$ , respectively. Now we apply the CP-like scheme, which maximizes the n-block upper bound  $\bar{C}_{fb,n}^{noisy}$  and  $\bar{C}_{fb,m}^{noisy}$ , as follows. Define a channel input process  $\{X_i\}_{i=1}^{n+m}$  as  $X_1^n = S_1^n + \mathbf{B}_n^*(W_1^n + V_1^n)$  and  $X_{n+1}^{n+m} = S_{n+1}^{n+m} + \mathbf{B}_n^*(W_{n+1}^{n+m} + V_{n+1}^{n+m})$ . Then

$$\begin{aligned} & n\bar{C}_{fb,n}^{noisy} + m\bar{C}_{fb,m}^{noisy} \\ &= I(X_1^n \rightarrow Y_1^n | V_1^{n-1}) + I(X_{n+1}^{n+m} \rightarrow Y_{n+1}^{n+m} | V_{n+1}^{n+m-1}) \\ &\stackrel{(a)}{=} I(S_1^n; Y_1^n | V_1^n) + I(S_{n+1}^{n+m}; Y_{n+1}^{n+m} | V_{n+1}^{n+m}) \\ &= h(S_1^n | V_1^n) + h(S_{n+1}^{n+m} | V_{n+1}^{n+m}) - h(S_1^n | Y_1^n, V_1^n) - h(S_{n+1}^{n+m} | Y_{n+1}^{n+m}, V_{n+1}^{n+m}) \\ &\stackrel{(b)}{=} h(S_1^n | V_1^{n+m}) + h(S_{n+1}^{n+m} | V_{n+1}^{n+m}) - h(S_1^n | Y_1^n, V_1^n) - h(S_{n+1}^{n+m} | Y_{n+1}^{n+m}, V_{n+1}^{n+m}) \\ &\stackrel{(c)}{=} h(S_1^{n+m} | V_1^{n+m}) - h(S_1^n | Y_1^n, V_1^n) - h(S_{n+1}^{n+m} | Y_{n+1}^{n+m}, V_{n+1}^{n+m}) \\ &\leq h(S_1^{n+m} | V_1^{n+m}) - h(S_1^{n+m} | Y_1^{n+m}, V_1^{n+m}) \\ &= I(S_1^{n+m}; Y_1^{n+m} | V_1^{n+m}) \\ &\stackrel{(d)}{=} h(Y_1^{n+m} | V_1^{n+m-1}) - h(W_1^{n+m}) \\ &\stackrel{(e)}{\leq} (n+m)\bar{C}_{fb,n+m}^{noisy} \end{aligned}$$

where line (a) and line (d) follow from Lemma 96. Line (b) and (c) follows from the block-wise white process construction. Line (e) follows from formula (5.11). Note that if  $\mathbf{K}_{s,n}^*$  or  $\mathbf{K}_{s,m}^*$



is singular and thus  $h(S_1^n|V_1^n)$  or  $h(S_{n+1}^{n+m}|V_{n+1}^{n+m})$  is ill-defined, we may apply the trick in the achievability proof of the Cover-Pombra theorem Cover and Pombra (1989), i.e., consider a sequence of nonsingular  $\mathbf{K}_{s,n}^*$  or  $\mathbf{K}_{s,m}^*$  that respectively achieves  $\bar{C}_{fb,n}^{noisy}$  or  $\bar{C}_{fb,m}^{noisy}$  in the limit. ■

### Proof of Lemma 60

**Proof.** We start with the supper-additive result as shown in the proof of Lemma 59,

$$n\bar{C}_{fb,n}^{noisy} + m\bar{C}_{fb,m}^{noisy} \leq (n+m)\bar{C}_{fb,n+m}^{noisy}$$

for all  $n, m \geq 1$ .

By taking  $m = n$ , we have  $\bar{C}_{fb,n}^{noisy} \leq \bar{C}_{fb,2n}^{noisy}$ . Then it is straightforward to obtain

$$3\bar{C}_{fb,n}^{noisy} \leq \bar{C}_{fb,n}^{noisy} + 2\bar{C}_{fb,2n}^{noisy} \leq 3\bar{C}_{fb,3n}^{noisy}$$

That is,  $\bar{C}_{fb,n}^{noisy} \leq \bar{C}_{fb,3n}^{noisy}$ . By repeating the above process, we have  $\bar{C}_{fb,n}^{noisy} \leq \bar{C}_{fb,kn}^{noisy}$  for all  $k \geq 1$ .

Next, define  $\Delta_k = \bar{C}_{fb,2^{k+1}n}^{noisy} - \bar{C}_{fb,2^kn}^{noisy}$ . According to the result above, we clearly have  $\Delta_k \geq 0$ . In addition,

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{k=1}^T \Delta_k &= \lim_{T \rightarrow \infty} \bar{C}_{fb,2^{T+1}n}^{noisy} - \bar{C}_{fb,2n}^{noisy} \\ &= \bar{C}_{fb}^{noisy} - \bar{C}_{fb,2n}^{noisy} < \infty. \end{aligned}$$

where the last line follows from Lemma 59. Therefore,  $\lim_{k \rightarrow \infty} \Delta_k = 0$ . ■

### Proof of Lemma 66

As  $\mathbf{C} \succeq \mathbf{B} \succ 0$  and  $\mathbf{A} \succeq 0$ , it is straightforward to have  $\det \mathbf{B}^{-1} \geq \det \mathbf{C}^{-1}$ , and thus

$$\det (\mathbf{B}^{-1} \mathbf{A} + \mathbf{I}) \geq \det (\mathbf{C}^{-1} \mathbf{A} + \mathbf{I}).$$

Then it is equivalent to have

$$\begin{aligned}
\frac{\det \mathbf{C}}{\det \mathbf{B}} &\geq \frac{\det \mathbf{C} \det (\mathbf{C}^{-1} \mathbf{A} + \mathbf{I})}{\det \mathbf{B} \det (\mathbf{B}^{-1} \mathbf{A} + \mathbf{I})} \\
\Leftrightarrow \frac{\det \mathbf{C}}{\det \mathbf{B}} &\geq \frac{\det (\mathbf{A} + \mathbf{C})}{\det (\mathbf{A} + \mathbf{B})} \\
\Leftrightarrow \frac{\det (\mathbf{A} + \mathbf{B})}{\det \mathbf{B}} &\geq \frac{\det (\mathbf{A} + \mathbf{C})}{\det \mathbf{C}} \\
\Leftrightarrow \log \frac{\det (\mathbf{A} + \mathbf{B})}{\det \mathbf{B}} &\geq \log \frac{\det (\mathbf{A} + \mathbf{C})}{\det \mathbf{C}}
\end{aligned}$$

### Proof of Lemma 69

The proof directly follows from the proof of Lemma 59 by carefully replacing the conditional directed information  $I(X_1^n \rightarrow Y_1^n | V_1^n)$  by the directed information  $I(X_1^n \rightarrow \tilde{Y}_1^n)$  on the new noiseless feedback Gaussian channel. For completeness and reader's convenience, we present the proof as follows. Remember that the n-block lower bound is obtained by solving n-block capacity of Gaussian channel with noiseless feedback where the Gaussian noise  $\tilde{W}_i = W_i + V_i$  (as shown in Fig. 5.5, right). According to Theorem 61, it is known that Cover-Pombra(CP) scheme can be applied to achieve this n-block noiseless feedback capacity. Then, we have

$$\begin{aligned}
I(S^n; \tilde{Y}^n) &= h(\tilde{Y}^n) - h(\tilde{Y}^n | S^n) \\
&= h(\tilde{Y}^n) - h(S^n + (\mathbf{I}_n + \mathbf{B}_n) \tilde{W}^n | S^n) \\
&\stackrel{(a)}{=} h(Y^n) - h((\mathbf{I}_n + \mathbf{B}_n) \tilde{W}^n) \\
&= h(\tilde{Y}^n) - h(\tilde{W}^n)
\end{aligned} \tag{C.4}$$

where line (a) follows from the fact that  $\tilde{W}^n = W^n + V^n$ , and random variables  $(S^n, W^n, V^n)$  are mutually independent.

Now we fix  $n$  and assume  $(\mathbf{K}_{s,n}^*, \mathbf{B}_n^*)$  achieves  $\underline{C}_{fb,n}^{noisy}$ . Similarly, fix  $m$  and assume  $(\mathbf{K}_{s,m}^*, \mathbf{B}_m^*)$  achieves  $\underline{C}_{fb,m}^{noisy}$ . Consider a process  $\{S_i\}_{i=1}^{n+m}$  that is independent of  $\{W_i\}_{i=1}^\infty$  and  $\{V_i\}_{i=1}^\infty$ , and block-wise white with  $\{S_i\}_{i=1}^n \sim \mathbb{N}_n(0, \mathbf{K}_{s,n}^*)$  and  $\{S_i\}_{i=n+1}^{n+m} \sim \mathbb{N}_m(0, \mathbf{K}_{s,m}^*)$ , respectively. Now we apply the CP scheme, which maximizes the n-block lower bound  $\underline{C}_{fb,n}^{noisy}$  and  $\underline{C}_{fb,m}^{noisy}$ , as follows. Define a channel input process  $\{X_i\}_{i=1}^{n+m}$  as  $X_1^n = S_1^n + \mathbf{B}_n^* \tilde{W}^n$  and  $X_{n+1}^{n+m} = S_{n+1}^{n+m} + \mathbf{B}_n^* \tilde{W}_{n+1}^{n+m}$ . Start with Remark 63, we have

$$\begin{aligned}
& n\bar{C}_{fb,n}^{noisy} + m\bar{C}_{fb,m}^{noisy} \\
&= h(\tilde{Y}_1^n) - h(\tilde{W}_1^n) + h(\tilde{Y}_{n+1}^{n+m}) - h(\tilde{W}_{n+1}^{n+m}) \\
&\stackrel{(a)}{=} I(S^n; \tilde{Y}^n) - I(S_{n+1}^{n+m}; \tilde{Y}_{n+1}^{n+m}) \\
&= h(S_1^n) + h(S_{n+1}^{n+m}) - h(S_1^n | \tilde{Y}_1^n) - h(S_{n+1}^{n+m} | \tilde{Y}_{n+1}^{n+m}) \\
&\stackrel{(b)}{=} h(S_1^{n+m}) - h(S_1^n | \tilde{Y}_1^n) - h(S_{n+1}^{n+m} | \tilde{Y}_{n+1}^{n+m}) \\
&\leq h(S_1^{n+m}) - h(S_1^{n+m} | \tilde{Y}_1^{n+m}) \\
&= I(S_1^{n+m}; \tilde{Y}_1^{n+m}) \\
&\stackrel{(c)}{=} h(\tilde{Y}_1^{n+m}) - h(\tilde{W}_1^{n+m}) \\
&\leq (n+m)\bar{C}_{fb,n+m}^{noisy}
\end{aligned}$$

Line (a) and (c) follow from equality (C.4). Line (b) follows from the block-wise white process construction. Similarly, if  $\mathbf{K}_{s,n}^*$  or  $\mathbf{K}_{s,m}^*$  is singular and thus  $h(S_1^n | V_1^n)$  or  $h(S_{n+1}^{n+m} | V_{n+1}^{n+m})$  is ill-defined, we may apply the trick in the achievability proof of the Cover-Pombra theorem Cover and Pombra (1989), i.e., consider a sequence of nonsingular  $\mathbf{K}_{s,n}^*$  or  $\mathbf{K}_{s,m}^*$  that respectively achieves  $\bar{C}_{fb,n}^{noisy}$  or  $\bar{C}_{fb,m}^{noisy}$  in the limit. According to Lemma 97, the property of super-additive sequence, the proof is complete.

## APPENDIX D. PROOFS OF RESULTS IN CHAPTER 6

### Proof of Lemma 77

Let  $\mathcal{A} = \{(m, z^n, y^n) : P_{M, Z^n, Y^n}(m|z^n, y^n) \leq P_{m, y^n}(m|y^n) \exp(-n\delta)\}$ . For every  $\delta > 0$ ,

$$\begin{aligned}
 & Pr\left\{\frac{1}{n} \log \frac{P_{M, Z^n, Y^n}(M|Z^n, Y^n)}{P_{M, Y^n}(M|Y^n)} \leq -\delta\right\} \\
 &= \sum_{\mathcal{A}} P_{M, Z^n, Y^n}(m, z^n, y^n) \\
 &= \sum_{\mathcal{A}} P_{M|Z^n, Y^n}(m|z^n, y^n) P_{Z^n, Y^n}(z^n, y^n) \\
 &\leq \sum_{\mathcal{A}} P_{M|Y^n}(m|y^n) P_{Z^n, Y^n}(z^n, y^n) \exp(-n\delta) \\
 &\leq \sum_{m, z^n, y^n} P_{M|Y^n}(m|y^n) P_{Z^n, Y^n}(z^n, y^n) \exp(-n\delta) \\
 &\leq \exp(-n\delta)
 \end{aligned}$$

The probability goes to zero as  $n \rightarrow \infty$ . Hence, we must have  $\underline{I}(M, Z^{n-1}|Y^n) \geq 0$ . By Lemma 71 and Fano's inequality,

$$\underline{I}(M, Z^{n-1}|Y^n) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} I(M, Z^{n-1}|Y^n) \leq \liminf_{n \rightarrow \infty} \delta_n.$$

where  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$ .

## APPENDIX E. PROOFS OF RESULTS IN CHAPTER 7

### Proof of Theorem 85

*Converse:* We firstly have

$$\begin{aligned}
 i(X^n \rightarrow Y^n || Z^{n-1}) &= \sum_{i=1}^n \log \frac{p(Y_i | X^i, Y^{i-1}, Z^{i-1})}{p(Y_i | Y^{i-1}, Z^{i-1})} \\
 &= \sum_{i=1}^n \log \frac{p(X_i + W_i | X^i, Y^{i-1}, W^{i-1}, Z^{i-1})}{p(Y_i | Y^{i-1}, Z^{i-1})} \\
 &\stackrel{(a)}{=} \sum_{i=1}^n \log \frac{p(W_i | W^{i-1})}{p(Y_i | Y^{i-1}, V^{i-1})} \\
 &\stackrel{(b)}{=} \sum_{i=1}^n \log \frac{p(W_i | W^{i-1})}{p(Y_i | Y^{i-1}, V^n)}
 \end{aligned}$$

where (a) follows from  $Z^{i-1} = [g_1(Y^1) + V_1, g_2(Y^2) + V_2, \dots, g_{i-1}(Y^{i-1}) + V_{i-1}]^T$  ( $g_i$  represents the operation of the feedback transmitter  $\mathbb{E}_2$ ). Line (b) follows from the Markov chain  $V_i^n - (Y^{i-1}, V^{i-1}) - Y_i$ . Therefore, we have  $I(X^n \rightarrow Y^n || Z^{n-1}) = h(Y^n | V^n) - h(W^n)$ . Now, consider a sequence of  $(n, 2^{nR_n})$  channel codes. By Fano's inequality, we have

$$\begin{aligned}
 nR_n &= H(M) \\
 &= H(M | Y^n, Z^{n-1}) + I(M; (Y^n, Z^{n-1})) \\
 &\stackrel{(a)}{=} I(X^n \rightarrow Y^n || Z^{n-1}) + n\delta_n \\
 &= h(Y^n | V^n) - h(W^n) + n\delta_n \\
 &\leq \max_{X^n} h(Y^n | V^n) - h(W^n) + n\delta_n
 \end{aligned}$$

where  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$ . Line (a) follows from Theorem 82. Now, it is well known that  $Y^n | V^n$  should be Gaussian to maximize  $h(Y^n | V^n)$ . Thus  $Y^n$  should be Gaussian as  $V^n$  is assumed to be Gaussian. Further we have  $Y^n = X^n + W^n$ , then  $X^n$  must be Gaussian.

Because  $X^n$  causally depends on the noises  $W^n$  and  $V^n$ , without loss of generality, the

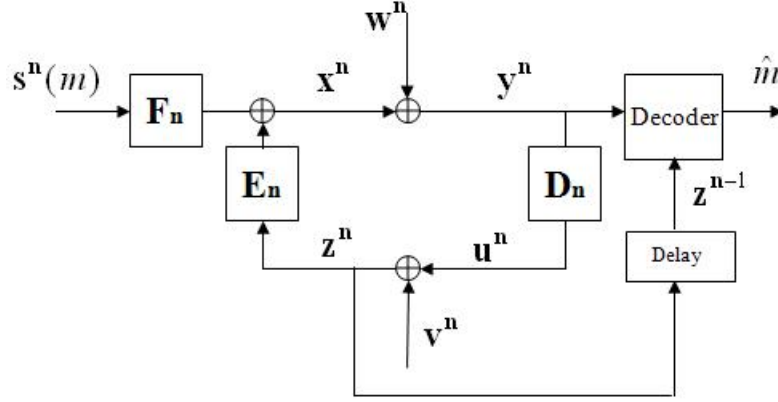


Figure E.1 Vector representation of the coding scheme (E.1) where  $\mathbf{E}_n = \mathbf{B}_n(\mathbf{I}_n + \mathbf{D}_n\mathbf{B}_n)^{-1}$  and  $\mathbf{F}_n = (\mathbf{I}_n + \mathbf{B}_n\mathbf{D}_n)^{-1}$ . Note that  $\mathbf{E}_n$  is strictly triangular, capturing the one-step feedback delay.

Gaussian channel inputs  $X^n$  can be constructed as

$$X^n = S^n + \mathbf{B}_n\mathbf{D}_nW^n + \mathbf{B}_nV^n \quad (\text{E.1})$$

where  $S^n \sim N(0, \mathbf{K}_{s,n})$  is the message information vector.  $\mathbf{B}_n$  is an  $n \times n$  strictly lower triangular linear matrix, capturing the one-step delay in the feedback channel.  $\mathbf{D}_n$  is an  $n \times n$  lower triangular linear matrix. Remark that random variables  $S^n, V^n, W^n$  are independent. This linear coding scheme is specifically presented in Fig.E.1.

By applying Lemma 3 and putting above formulas together, we have  $R_n \leq C_{FB,n} + \delta_n$ . The converse thus is proved.

*Achievability:* As did in Cover and Pombra (1989), the achievability proof follows from the standard random coding technique and AEP. Consider the coding scheme (E.1) where  $\mathbf{K}_{s,n}$ ,  $\mathbf{B}_n$  and  $\mathbf{D}_n$  achieve  $C_{FB,n}$ . Since  $S^n$  is the message information vector and only determined by message index  $M$ , according to Theorem 82, it is easy to show

$$I(S^n; (Y^n, Z^{n-1})) = I(X^n \rightarrow Y^n | Z^{n-1}) = h(Y^n | V^n) - h(W^n). \quad (\text{E.2})$$

Let  $\tilde{Y}^n = (Y^n, Z^{n-1})$  and  $(S^n, \tilde{Y}^n)$  be jointly distributed with density  $f(S^n, \tilde{Y}^n)$ . Then the

set  $\mathcal{A}_\epsilon^n$  of jointly  $\epsilon$ -typical  $(S^n, \tilde{Y}^n)$  is defined by

$$\mathcal{A}_\epsilon^n = \left\{ \left| -\frac{1}{n}f(S^n) - \frac{1}{n}h(S^n) \right| \leq \epsilon, \left| -\frac{1}{n}f(\tilde{Y}^n) - \frac{1}{n}h(\tilde{Y}^n) \right| \leq \epsilon \right. \\ \left. \left| -\frac{1}{n}f(S^n, \tilde{Y}^n) - \frac{1}{n}h(S^n, \tilde{Y}^n) \right| \leq \epsilon \right\}.$$

By Lemma 6 in Cover and Pombra (1989), we have the volume of  $\mathcal{A}_\epsilon^n$  as

$$V(\mathcal{A}_\epsilon^n) \leq 2^{(h(S^n, \tilde{Y}^n) + n\epsilon)}. \quad (\text{E.3})$$

Let  $S^n(1), S^n(2), \dots, S^n(2^{nR})$  be i.i.d. vectors drawn according to  $N(0, \mathbf{K}_{s,n})$ <sup>1</sup>. To send message  $M$ , the transmitter sends out  $X^n = S^n(M) + \mathbf{B}_n \mathbf{D}_n W^n + \mathbf{B}_n V^n$ . The receiver then declares  $\hat{M}$  was sent if  $(S^n(\hat{M}), \tilde{Y}^n)$  is the only  $\epsilon$ -typical pair. If there is no typical pair or more than one such or  $\hat{M} \neq M$ , an error is declared. Assume  $M = 1$  is sent and define

$$\mathcal{E}_i : (S^n(i), \tilde{Y}^n) \in \mathcal{A}_\epsilon^n.$$

and  $\mathcal{E}_i^c$ , the complement of  $\mathcal{E}_i$ . Then we have

$$P_e^{(n)} \leq Pr(\mathcal{E}_1^c | M = 1) + 2^{nR} Pr(\mathcal{E}_2 | M = 1).$$

By AEP, the first term converges to zero as  $n \rightarrow \infty$ . Next,

$$\begin{aligned} Pr(\mathcal{E}_2 | M = 1) &= \int_{(s^n, \tilde{y}^n) \in \mathcal{A}_\epsilon^n} f(s^n) f(\tilde{y}^n) ds^n d\tilde{y}^n \\ &\stackrel{(a)}{\leq} 2^{(h(S^n, \tilde{Y}^n) - h(S^n) - h(\tilde{Y}^n) + 3n\epsilon)} \\ &= 2^{-I(S^n; (Y^n, Z^{n-1})) + 3n\epsilon} \\ &\stackrel{(b)}{=} 2^{-(h(Y^n | V^n) - h(W^n)) + 3n\epsilon} \\ &= 2^{(-nC_{FB,n} + 3n\epsilon)} \end{aligned}$$

where (a) follows from (E.3) and (b) follows from equation (E.2). Putting above together, we conclude that, for  $R < C_{FB,n} - 3\epsilon$ , there exists a sequence of  $(n, 2^{nR})$  channel codes with  $P_e^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$ .

---

<sup>1</sup>In this proof,  $\mathbf{K}_{s,n}$  is assumed to be nonsingular such that the AEP will apply. We can use the trick in Cover and Pombra (1989) to deal with the nonsingular case.

### Proof of Corollary 87

We show that the optimization problem (7.4) can be casted into the bilinear form (7.5).

Let

$$\begin{aligned}\mathbf{Y}_n &= \mathbf{K}_{s,n} + (\mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n) \mathbf{K}_{w,n} (\mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n)^T \\ &= \mathbf{K}_{s,n} + \mathbf{B}_n \mathbf{D}_n \mathbf{K}_{w,n} \mathbf{D}_n^T \mathbf{B}_n^T + \mathbf{K}_{w,n} \mathbf{D}_n^T \mathbf{B}_n^T + \mathbf{B}_n \mathbf{D}_n \mathbf{K}_{w,n} + \mathbf{K}_{w,n}\end{aligned}$$

Then the power constraints and  $\mathbf{K}_{s,n} \geq 0$  can be alternatively expressed as

$$\begin{aligned}tr(\mathbf{Y}_n - \mathbf{B}_n \mathbf{D}_n \mathbf{K}_{w,n} - \mathbf{K}_{w,n} \mathbf{D}_n^T \mathbf{B}_n^T - \mathbf{K}_{w,n} + \mathbf{B}_n \mathbf{K}_{v,n} \mathbf{B}_n^T) &\leq nP_1, \\ tr(\mathbf{D}_n \mathbf{Y}_n \mathbf{D}_n^T + \mathbf{D}_n \mathbf{B}_n \mathbf{K}_{v,n} \mathbf{B}_n^T \mathbf{D}_n^T) &\leq nP_2, \\ \mathbf{Y}_n - (\mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n) \mathbf{K}_{w,n} (\mathbf{I}_n + \mathbf{B}_n \mathbf{D}_n)^T &\geq 0.\end{aligned}$$

By introducing dummy matrices  $\mathbf{H}_n$  and  $\mathbf{Z}_n$  and then applying Schur complement decomposition, it is straightforward to obtain the bilinear optimization (7.5).

### Proof of Theorem 90

A). (*Power Constraint*) We show that the coding structure  $\mathcal{C}$  satisfies the power constraints.

First of all, we derive two necessary results as follows. Let

$$g(x) = (1 + \alpha^2)x^3 + 2\alpha x^2 - (1 + \alpha^2 + P_1)x - 2\alpha. \quad (\text{E.4})$$

It is straightforward to obtain  $g(1) = -P_1$ ,  $g(-1) = P_1$ ,  $g(+\infty) = +\infty$ ,  $g(-\infty) = -\infty$  and

$$g(-r_i) = 4\alpha(r_i^2 - 1),$$

where  $r_i$  ( $i = 1, 2, 3$ ) is the root of polynomial  $g(x) = 0$ . After some inference, we conclude that all the three roots are real and  $r_1 < -1 < r_2 < 1 < r_3$ . Furthermore, we have

$$\begin{aligned}|r_1| &< r_3 & \text{for } \alpha < 0, \\ |r_1| &> r_3 & \text{for } \alpha > 0, \\ |r_1| &= r_3 & \text{for } \alpha = 0.\end{aligned}$$



Let  $A(\xi_i = 1) = -r$  ( $i = 0, 1, \dots, n$ ) where  $r$  is the real root with the largest absolute value.

Then we have

$$\alpha A^{-1}(\xi_i = 1) \geq 0. \quad (\text{Necessary Result 1}) \quad (\text{E.5})$$

Furthermore, we can derive that

$$\alpha B(\xi_i = 1) \leq 0. \quad (\text{Necessary Result 2}) \quad (\text{E.6})$$

Next, since  $A(\xi_i = 1) = -r$  is a real root of the polynomial  $g(x) = 0$ , we clearly have  $(r^2 - 1)g(r) = 0$ . After some algebra, we have

$$B^2(\xi_i = 1) = \left(\frac{1}{r} - r\right)^2 = \frac{P_1 - r^{-2}P_1}{2\alpha r^{-1} + 1 + \alpha^2}. \quad (\text{Necessary Result 3}) \quad (\text{E.7})$$

Now, we are ready to show that the coding structure  $\mathcal{C}$  satisfies the power constraints  $P_1$  and  $P_2$ .

$$\begin{aligned} X_n &= A(\xi_{n-1})X_{n-1} + \frac{1}{g}B(\xi_{n-1})\xi_{n-1}U_{n-1} \\ &= A(\xi_{n-1})X_{n-1} + \frac{1}{g}B(\xi_{n-1})\xi_{n-1}gY_{n-1} \\ &= A(\xi_{n-1})X_{n-1} + B(\xi_{n-1})\xi_{n-1}Y_{n-1} \\ &= (A(\xi_{n-1}) + B(\xi_{n-1})\xi_{n-1})X_{n-1} + B(\xi_{n-1})\xi_{n-1}\tilde{U}_{n-1} + \alpha B(\xi_{n-1})\xi_{n-1}\tilde{U}_{n-2}. \end{aligned}$$

Taking  $\xi_{n-1} = 0$  and  $1$ , respectively, we can obtain

$$X_n = A^{-1}(\xi_{n-1})X_{n-1} + B(\xi_{n-1})\tilde{U}_{n-1} + \alpha B(\xi_{n-1})\tilde{U}_{n-2}.$$

Then we have

$$\begin{aligned} &\mathbb{E}[X_n\tilde{U}_{n-1}] \\ &= \mathbb{E}[(A(\xi_{n-1})^{-1}X_{n-1} + B(\xi_{n-1})\tilde{U}_{n-1} + \alpha B(\xi_{n-1})\tilde{U}_{n-2})\tilde{U}_{n-1}] \\ &= B(\xi_{n-1})\mathbb{E}[\tilde{U}_{n-1}^2] \\ &= B(\xi_{n-1}) \end{aligned}$$

The average transmission power at time  $n$  is

$$\begin{aligned}
& \mathbb{E}[X_n^2] \\
&= A^{-2}(\xi_{n-1})\mathbb{E}[X_{n-1}^2] + 2\alpha A^{-1}(\xi_{n-1})B(\xi_{n-1})B(\xi_{n-2}) + B^2(\xi_{n-1}) + \alpha^2 B^2(\xi_{n-1}) \\
&\leq A^{-2}(\xi_{n-1})\mathbb{E}[X_{n-1}^2] + 2\alpha A^{-1}(\xi_{n-1})B^2(\xi_{n-1}) + B^2(\xi_{n-1}) + \alpha^2 B^2(\xi_{n-1}) \\
&= A^{-2}(\xi_{n-1})\mathbb{E}[X_{n-1}^2] - A^{-2}(\xi_{n-1})P_1 + P_1.
\end{aligned} \tag{E.8}$$

where the last two lines follow from the necessary results 1 and 3, respectively. Thus,

$$\begin{aligned}
\mathbb{E}[X_n^2] - P_1 &\leq A^{-2}(\xi_{n-1})(\mathbb{E}[X_{n-1}^2] - P_1) \\
&\leq \prod_{j=0}^{n-1} A^{-2}(\xi_j)(\mathbb{E}[X_0^2] - P_1)
\end{aligned}$$

Since  $A^{-2}(1) = r^{-2} \leq 1$ , we have  $\mathbb{E}[X_n^2]$  converges to  $P_1$  almost surely<sup>2</sup> as  $n \rightarrow \infty$ . This implies the time average of  $\mathbb{E}[X_n^2]$  is  $P_1$  as  $n \rightarrow \infty$ .

Now, we check the power constraint  $P_2$ . That is,

$$\begin{aligned}
\mathbb{E}[U_n^2] &= g^2 \mathbb{E}[Y_n^2] \\
&= g^2 \mathbb{E}[(X_n + \tilde{U}_n + \alpha \tilde{U}_{n-1})^2] \\
&= g^2 (\mathbb{E}[X_n^2] + 2\alpha B(\xi_{n-1}) + 1 + \alpha^2) \\
&\leq g^2 (\mathbb{E}[X_n^2] + 1 + \alpha^2)
\end{aligned} \tag{E.9}$$

where the last inequality follows from the necessary result 2 the definition of  $B(\xi)$ . Therefore,

$$\mathbb{E}[U_n^2] \stackrel{a.s.}{\leq} g^2 (P_1 + 1 + \alpha^2) \leq P_2, \tag{E.10}$$

which implies the time average of  $\mathbb{E}[U_n^2]$  is less than  $P_2$ .

B) (*Decoding Error and Transmission Rate*) We now investigate the probability of decoding error and the transmission rate. According to the coding strategy  $\mathcal{C}$ , it is easy to obtain

$$X_{n-T+1} - \tilde{X}_n = \prod_{j=0}^{n-T} A(\xi_j)(X_0 - \tilde{X}_{T-1}).$$

---

<sup>2</sup>Note  $\{\xi_j\}_0^{n-1}$  is a Bernoulli process with probability  $1 - e$ . By law of large number, we have for any  $\epsilon > 0$ ,  $Pr(|\lim_{n \rightarrow \infty} \frac{\sum_{j=0}^{n-1} \xi_j}{n} - (1 - e)| < \epsilon) = 1$ . Such a sequence of  $\{\xi_j\}_0^{n-1}$  is defined as a “typical” sequence, which then implies  $Pr(|\lim_{n \rightarrow \infty} \prod_{j=0}^{n-1} a^{-2}(\xi_j) - 0| < \epsilon) = 1$ . We herein are concerning typical sequences only and use “almost sure” to capture events happening with probability 1.

Then, with  $\tilde{X}_{T-1} = 0$ , we have

$$\hat{X}_{0,n} = \prod_{j=0}^{n-T} A(\xi_j)^{-1} \tilde{X}_n = \prod_{j=0}^{n-T} A(\xi_j)^{-1} X_{n-T+1} - X_0.$$

For large  $n$ ,  $\hat{X}_{0,n}$  has a distribution with mean  $-X_0$  and bounded variance

$$\sigma_n^2 = \prod_{j=0}^{n-T} A(\xi_j)^{-2} \mathbb{E}[X_{n-T+1}^2] \stackrel{a.s.}{=} P_1 r^{-2(1-e)(n-T+1)}.$$

Now, we equally partition the interval  $(-\sqrt{P_1}, \sqrt{P_1})$  into  $M_n = \sigma_n^{-(1-\epsilon)}$  ( $\epsilon > 0$ ) where the center of each segment represents a message to be transmit. By Chebyshev's inequality, we have the probability of error (i.e.  $X_0$  and  $\hat{X}_{0,n}$  locate in different segments) as

$$P_{e,n} = \Pr\left(|\hat{X}_{0,n} - X_0| \geq \frac{\sqrt{P_1}}{\sigma_{n-1}^{-(1-\epsilon)}}\right) \stackrel{a.s.}{\leq} P_1^{(\epsilon-1)} r^{-2(1-e)\epsilon(n-T+1)}.$$

This implies that the error probability decays at least exponentially. Besides, we see that the error probability is also affected by the finite delay  $T$ . Finally, we have the transmission rate

$$\begin{aligned} R &= \lim_{n \rightarrow \infty} \frac{\log M_n}{n} \\ &= \lim_{n \rightarrow \infty} \frac{-(1-\epsilon) \log(\prod_{j=0}^{n-T} A(\xi_j)^{-2} \mathbb{E}[X_{n-T+1}^2])}{2n} \\ &\stackrel{a.s.}{=} \lim_{n \rightarrow \infty} \frac{-(1-\epsilon)}{2n} (-(1-e)(n-T+1) \log(A^2(1) + \log P)) \\ &= (1-\epsilon)(1-e) \log(|r|). \end{aligned} \tag{E.11}$$

## BIBLIOGRAPHY

- Agrawal, M. and Love, D. J. (2011). An iteratively optimized linear coding scheme for correlated gaussian channels with noisy feedback. *Allerton Conference in Communication, control and computing*.
- Alajaji, F. (1994). Feedback does not increase the capacity of discrete channels with additive noise. *IEEE Transactions on Information Theory*, 40(4):1147–1157.
- Alajaji, F. and Fuja, T. (1994). Effect of feedback on the capacity of discrete additive channels with memory. *Proc. of the IEEE International Symposium on Information Theory*.
- Ardestanizadeh, E. and Franceschetti, M. (2012). Control-theoretic approach to communication with feedback. *IEEE Transactions on Automatic Control*, 57(10):2576–2587.
- Ardestanizadeh, E., PaoloMinero, and Franceschetti, M. (2012). Lqg control approach to gaussian broadcast channels with feedback. *IEEE Transactions on Information Theory*, 58(8):5267–5278.
- Baillieul, J. (2002). Feedback coding for information-based control: operating near the data-rate limit. *Proc. of the IEEE Conf on Decision and Control*, pages 3229–3236.
- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge university, Cambridge.
- Burnashev, M. V. and Yamamoto, H. (2008). On bsc, nosiy feedback and three message. *In Proc. IEEE International Symposium on Information Theory*, pages 6–11.
- Butman, S. A. (1969). A general formulation of linear feedback communication systems with solutions. *IEEE Transaction on Information Theory*, 15(3):392–400.

- Butman, S. A. (1976). Linear feedback rate bounds for regressive channels. *IEEE Transaction on Information Theory*, 22(3):363–366.
- Caire, G. and Shamai(Shitz), S. (1999). On the capacity of some channels with channel state information. *IEEE Transactions on Information Theory*, 45(6):2007–2019.
- Chance, Z. and Love, D. J. (2010). A noisy feedback encoding scheme for the gaussian channel. *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 3482–3485.
- Chance, Z. and Love, D. J. (2011a). Concatenated coding for the awgn channel with noisy feedback. *IEEE Transactions on Information Theory*, 57(10):6633–6649.
- Chance, Z. and Love, D. J. (2011b). Concatenated coding for the awgn channel with noisy feedback. *IEEE Transactions on Information Theory*, 57(10):6633–6649.
- Chen, H. W. and Yanagi, K. (1999). Refinements of the half-bit and factor-of-two bounds for capacity in gaussian channel with feedback. *IEEE Transaction on Information Theory*, 45(1):319–325.
- Chen, J. and Berger, T. (2005). The capacity of finite-state markov channels with feedback. *IEEE Transactions on Information Theory*, 51(3):780–789.
- Cover, T. M. and Pombra, S. (1989). Gaussian feedback capacity. *IEEE Transactions on Information Theory*, 35(1):37–43.
- Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory, 2nd edition*. Wiley, New York.
- Dembo, A. (1989). On gaussian feedback capacity. *IEEE Transaction on Information Theory*, 35(5):1072–1076.
- Draper, S. C. and Sahai, A. (2006a). Beating burnashev in delay with noisy feedback. *Allerton Conference in Communication, control and computing*.
- Draper, S. C. and Sahai, A. (2006b). Noisy feedback improves communication reliability. *In Proc. IEEE International Symposium on Information Theory*, pages 69–73.

- Draper, S. C. and Sahai, A. (2008). Variable-length channel coding with noisy feedback. *European Transactions on Telecommunications*, 19:355–370.
- Ebert, P. (1970). The capacity of the gaussian channel with feedback. *Bell Syst. Tech. J.*, pages 1705–1712.
- Elia, N. (2002). Stabilization of systems with erasure actuation and sensory channels. *Allerton Conference on Communication, Control, and Computing*, pages 315–324.
- Elia, N. (2003). Feedback stabilization in the presence of fading channels. *American Control Conference*, pages 4438–4443.
- Elia, N. (2004). When bode meets shannon: control-oriented feedback communication schemes. *IEEE Transactions on Automatic Control*, 49(9):1477–1488.
- Elia, N. (2005). The information cost of loop shaping over gaussian channels. *Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation Intelligent Control*.
- Elia, N. and Mitter, S. K. (2001). Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 46(9):1384–1400.
- Elias, P. (1956). Channel capacity without coding. *MIT Research Lab. of Electronics, Cambridge, MA, Quarterly Progr. Rep.*
- Elias, P. (1967). Networks of gaussian channels with applications to feedback systems. *IEEE Transactions on Information Theory*, 13:493–501.
- Gallager, R. G. (1968). *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., United States of America.
- Gastpar, M. and Kramer, G. (2006). On cooperation via noisy feedback. *International Zurich Seminar on Communications*, pages 146–149.
- Han, T. and Verdú, S. (1993). Approximation theory of output statistics. *IEEE Transactions on Information Theory*, 39(3):752–772.

- Han, T. S. (2003). *Information-Spectrum Methods in Information Theory*. Springer, Berlin.
- Ihara, S. (1980). On the capacity of the continuous time gaussian channel with feedback. *J. Multivariate Anal.*, 10(3):319–331.
- Ihara, S. (1988). Capacity of discrete time gaussian channel with and without feedback. *Mem. Fac. Sci. Kochi Univ. Ser. A Math.*, 9(3):21–36.
- Ihara, S. (1990). Capacity of mismatched gaussian channels with and without feedback. *Probab. Theory Rel. Fields*, 84(4):453–471.
- Ihara, S. (1994). Coding theorems for a continuous-time gaussian channel with feedback. *IEEE Transactions on Information Theory*, 40:2014–2045.
- Kim, Y. H. (2008). A coding theorem for a class of stationary channels with feedback. *IEEE Transactions on Information Theory*, 54(4):1488–1499.
- Kim, Y. H. (2010). Feedback capacity of stationary gaussian channels. *IEEE Transactions on Information Theory*, 56(1):57–85.
- Kim, Y. H., Lapidoth, A., and Weissman, T. (2007). The gaussian channel with noisy feedback. *ISIT*, pages 1416–1420.
- Kramer, G. (2002a). Directed information for channels with feedback. *Swiss Federal Institute of Technology Zurich, Ph.D dissertation*, 48:1423–1438.
- Kramer, G. (2002b). Feedback strategies for white gaussian interference networks. *IEEE Transaction on Information Theory*, 48:1423–1438.
- Kumar, U., Laneman, J. N., and Gupta, V. (2009). Noisy feedback schemes and rate-error tradeoffs from stochastic approximation. *IEEE International Symposium on Information Theory*, pages 1258–1262.
- Lapidoth, A. and Wigger, M. (2010). On the awgn mac with imperfect feedback. *IEEE Transactions on Information Theory*, 56(11):5432–5476.

- Lavenberg, S. S. (1971). Repetitive signaling using a noisy feedback channel. *IEEE Transactions on Information Theory*, IT-17(3):269–278.
- Li, C. and Elia, N. (2011a). Bounds on the achievable rate of noisy feedback gaussian channels under linear feedback coding scheme. *IEEE International Symposium on Information Theory*, pages 169–173.
- Li, C. and Elia, N. (2011b). The information flow and capacity of channels with noisy feedback. *arXiv preprint arXiv:1108.2815*.
- Li, C. and Elia, N. (2011c). The information theoretic characterization of the capacity of channels with noisy feedback. *IEEE International Symposium on Information Theory*, pages 174–178.
- Li, C. and Elia, N. (2011d). Upper bound on the capacity of gaussian channels with noisy feedback. *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 84–89.
- Li, C. and Elia, N. (2012). Noisy feedback communications with side information at the decoder. *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1856–1863.
- Liu, J. L. (2006). Fundamental limits in gaussian channels with feedback: confluence of communication, estimation, and control. *ISU Ph.D. dissertation*.
- Liu, J. L. and Elia, N. (2005). Writing on dirty paper with feedback. *Communications on Information and Systems*, 5(4):401–422.
- Liu, J. L. and Elia, N. (2006). Convergence of fundamental limitations in information, estimation, and control. *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 5609–5614.
- Liu, J. L., Elia, N., and Tatikonda, S. (2004a). Capacity achieving feedback scheme for flat fading channels with channel state information. *American Control Conference*, pages 3593–3597.



- Liu, J. L., Elia, N., and Tatikonda, S. (2004b). Capacity achieving feedback scheme for markov channels with channel state information. *in Proceedings IEEE International Symposium on Information Theory*, pages 71–75.
- Lucky, R. W. (1973). A survey of the communication theory literature: 1968c1973. *IEEE Transactions on Information Theory*, 19:725–739.
- Martins, N. and Dahleh, M. A. (2008). Feedback control in the presence of noisy channels: Bode-like fundamental limitations of performance. *IEEE Transactions on Automatic Control*, 53(7):1604 – 1615.
- Martins, N. C. and Weissman, T. (Sep.2008). Coding for additive white noise channels with feedback corrupted by quantization or bounded noise. *IEEE Transactions on Information Theory*, 54(9):4274–4282.
- Massey, J. L. (1990). Causality, feedback and directed information. *IEEE International Symposium on Information Theory*.
- Matveev, A. S. and Savkin, A. V. (2001). Optimal control of networked systems via asynchronous communication channels with irregular delays. *Proceedings of the 40th IEEE Conference on Decision and Control*, 3:2327 –2332.
- Nair, G. N. and Evans, R. J. (2000). Communication-limited stabilization of linear systems. *Proceedings of the 39th IEEE Conference on Decision and Control*, pages 1005 –1010.
- Omura, J. K. (1968). Optimum linear transmission of analog data for channels with feedback. *IEEE Transactions on Information Theory*, IT-14(1):38–43.
- Ozarow, L. (1984). The capacity of the white noise gaussian multiple-access channels with feedback. *IEEE Transaction on Information Theory*, 30(4):623–629.
- Ozarow, L. H. (1990a). Random coding for additive gaussian channels with feedback. *IEEE Transactions on Information Theory*, 36(1):17–22.
- Ozarow, L. H. (1990b). Upper bounds on the capacity of gaussian channels with feedback. *IEEE Transactions on Information Theory*, 36(1):156–161.

- Ozarow, L. H. and Leung, C. S. K. (1984). An achievable region and an outer bound for the gaussian broadcast channel with feedback. *IEEE Transaction on Information Theory*, 30(4):667–671.
- Permuter, H. H., Kim, Y. H., and Weissman, T. (2011). Interpretations of directed information in portfolio theory, data compression, and hypothesis testing. *IEEE Transactions on Information Theory*, 57(6):3248–3259.
- Permuter, H. H., Weissman, T., and Goldsmith, A. J. (2009). Finite state channels with time-invariant deterministic feedback. *IEEE Transactions on Information Theory*, 55(2):644–662.
- Pombra, S. and Cover, T. (1994). Non-white gaussian multiple access channel with feedback. *IEEE Transaction on Information Theory*, 40(3):885–892.
- Pradhan, S. S. (2004). Source coding with feedforward: Gaussian sources. *Proc.of the IEEE International Symposium on Information Theory*, page 212.
- Sahai, A. (2006). The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication linkpart i: Scalar systems. *IEEE Transactions on Information Theory*, 52(8):3369–3395.
- Sahai, A. (Feb. 2001). Anytime information theory. *MIT Ph.D. dissertation*.
- Schalkwijk, J. P. (1968). Center of gravity information feedback. *IEEE Transactions on Information Theory*, 14:324–331.
- Schalkwijk, J. P. M. (1966). A coding scheme for additive noise channels with feedback II: bandlimited signals. *IEEE Transactions on Information Theory*, IT-12:183–189.
- Schalkwijk, J. P. M. and Kailath, T. (1966). A coding scheme for additive noise channels with feedback I: No bandwidth constraint. *IEEE Transactions on Information Theory*, IT-12:172–182.
- Sen, N., Alajaji, F., and Yuksel, S. (2011). Feedback capacity of a class of symmetric finite-state markov channels. *IEEE Transactions on Information Theory*, 57(7):4110–4122.

- Shahar-Doron, A. and Feder, M. (2004). On a capacity achieving scheme for the colored gaussian channel with feedback. *IEEE International Symposium on Information Theory*, page 74.
- Shannon, C. (1958). Channels with side information at the transmitter. *IBM J. Res. Develop.*, 2:289–293.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656.
- Shannon, C. E. (1949). Communication in the presence of noise. *Proc. IRE*, 37:10–21.
- Shayevitz, O. and Feder, M. (2011). Optimal feedback communication via posterior matching. *IEEE Transactions on Information Theory*, 57(3):1186 – 1222.
- Shrader, B. and Permuter, H. (2009). Feedback capacity of the compound channel. *IEEE Transactions on Information Theory*, 55(8):3629–3644.
- Tatikonda, S. (Aug. 2000). Control under communication constraints. *MIT Ph.D. dissertation*.
- Tatikonda, S. and Mitter, S. (2000). Control under communication constraints. *Allerton Conference on Communication, Control and Computing*.
- Tatikonda, S. and Mitter, S. (2009). The capacity of channels with feedback. *IEEE Transactions on Information Theory*, 55(1):323–349.
- Tiernan, J. C. and Schalkwijk, J. P. M. (1974). An upper bound to the capacity of the band limited gaussian autoregression channel with noiseless feedback. *IEEE Transaction on Information Theory*, 20(3):311–316.
- Vandenberghe, L., Boyd, S., and Wu, S. P. (1998). Determinant maximization with linear matrix inequality constraints. *SIAM J. Matrix Anal. Appl.*, 19(2):499–533.
- Venkataramanan, R. and Pradhan, S. (2007). Source coding with feedforward: Rate-distortion theorems and error exponents for a general source. *IEEE Transactions on Information Theory*, 53(6):2154–2179.

- Verdú, S. and Han, T. (1994). A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4):1147–1157.
- Viswanathan, H. (1999). Capacity of markov channels with receiver sci and delayed feedback. *IEEE Transactions on Information Theory*, 45(2):761–771.
- Wang, J. and Elia, N. (2011). A control perspective for centralized and distributed convex optimization. *Proceedings of the 50th IEEE Conference on Decision and Control*, pages 3800–3805.
- Weissman, T. and Merhav, N. (2003). On competitive prediction and its relation to rate-distortion theory. *IEEE Transactions on Information Theory*, 49(12):3185–3194.
- Witsenhausen, H. S. (1968). A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6(1):131–147.
- Wolfowitz, J. (1975). Signalling over a gaussian channel with feedback and autoregressive noise. *J. Appl. Probability*, 12(4):713–723.
- Wong, W. S. and Brockett, R. W. (1999). Systems with finite communication bandwidth constraints. ii. stabilization with limited information feedback. *IEEE Transactions on Automatic Control*, 44(5):1049–1053.
- Wyner, A. D. (Dec.1969). On digital communication over a discere-time gaussian channel with noisy feedback. *The Bell System Technical Journal*, pages 3173–3186.
- Yang, S., Kavcic, A., and Tatikonda, S. (2005). Feedback capacity of finite-state machine channels. *IEEE Transactions on Information Theory*, 51(3):799–810.
- Yang, S. H. (2004). The capacity of communication channels with memory. *Harvard, Ph.D. dissertation*.
- Yang, S. H. (2007). On the feedback capacity of power-constrained gaussian noise channels with memory. *IEEE Transactions on Information Theory*, 53(3):929–954.

- Yuksel, S. and Bassar, T. (2011). Control over noisy forward and reverse channels. *IEEE Transactions on Automatic Control*, 56(5):1014–1029.
- Zamir, R., Kochman, Y., and Erez, U. (2008). Achieving the gaussian rate-distortion function by prediction. *IEEE Transactions on Information Theory*, 54(7):3354–3364.
- Zhang, L. and Guo, D. (2011). Capacity of gaussian channels with duty cycle and power constraints. In *IEEE International Symposium on Information Theory (ISIT)*, pages 513–517.